Author:
Gaunt, Jess

Title:
A Viral Approach to Translatome Profiling of CA1 Neurons During Associative Recognition Memory Formation

A Viral Approach to Translatome Profiling of CA1 Neurons During Associative Recognition Memory Formation

Jessica Ruth Gaunt

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Health Sciences, Bristol Medical School.

September 2017

Word Count: ~52,000

Abstract

Associative recognition memory enables judgements of whether configurations of stimuli have been previously encountered and these memories require de novo protein synthesis in the CA1 subregion of the hippocampus for their consolidation. To investigate associative recognition memory in the present study, the paired viewing procedure (Zhu et al, 1996) was used to present novel and familiar arrangements of images using a within-subjects design. A previous report of increased Fos expression in the novel compared to the familiar condition in CA1 following paired viewing of arrangements was replicated.

As there is increased sensitivity to subtle changes in gene expression in the translatome compared to the transcriptome, translating ribosome affinity purification (TRAP; Heiman et al, 2008) was used to extract messenger RNAs engaged with the ribosome by tagging the ribosomal protein L10a with EGFP. In this thesis, viral vectors expressing the EGFP-L10a transgene were developed to target CA1 pyramidal neurons.

Combining TRAP and paired viewing, a study profiling the translatome of rat CA1 neurons over time following paired viewing was conducted, aiming to further understanding of the genes and regulatory networks involved in associative recognition memory. As gene expression is also induced in response to behavioural conditions not requiring learning, a no-image control condition was included with matched timepoints. TRAP-generated RNA samples were sequenced and submitted to a range of bioinformatic analyses to identify genes that were differentially expressed between novel and familiar conditions and over time. There was strong evidence of enriched protein-protein interactions between differentially expressed genes (DEGs) and potential regulators of the DEG network were identified. This first translatome profiling study of associative recognition memory sheds light on the dynamics of gene expression during learning and identifies promising candidate genes and molecular functions of interest for follow-up studies.

Acknowledgements

I would firstly like to thank my supervisors, James Uney, Clea Warburton, and Lucia Marucci, for all their advice and support during my PhD. I would particularly like to thank Helen Scott for her patient guidance in the lab and her assistance with TRAP experiments. I would also like to thank collaborators Steven Sheardown for his advice in setting up the TRAP protocol, Younbok Lee and Doyoung Lee for producing the AAV, and Mark Rogers and Tom Batstone for advice on bioinformatics.

I would also like to thank the members of the Bashir-Warburton and Uney-Wong labs, as well as the other present and former students on the Neural Dynamics course, for their friendship and support over the past four years. I am also grateful to the Neural Dynamics programme directors and the Wellcome Trust for giving me this opportunity.

I am also grateful to the staff of the Bristol Animal Services Unit for all of the work they do that enabled these experiments and would particularly like to thank Gareth Gough and Jo Roe for going above and beyond. I would also like to thank Simon Lishman of the IT department and Tom Richardson of the Electronics workshop for all their help updating the behavioural equipment.

Finally, I would like to thank my friends and family for their endless love and support on this journey.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED:
..... DATE:.....

Table of Contents

Chapter 1 Introduction 1

1.1 Memory and the brain 1

1.1.1 Types of memory and memory processes 1

1.1.2 Neural pathways for memory and regional functions 2

1.1.3 Recognition memory 6

1.2 Genes and cellular processes in the control of synaptic plasticity and memory formation 13

1.2.1 Types of plasticity 14

1.2.2 Regulation of gene expression for plasticity 16

1.2.3 Differential gene expression and plasticity 19

1.3 Profiling genome-wide gene expression 23

1.3.1 Identifying mRNA transcripts 23

1.3.2 Translatome profiling 25

1.4 Modification of gene expression in-vivo 30

1.4.1 Transgenic organisms 30

1.4.2 Viruses 31

1.4.3 Summary 35

1.5 Present work 37

Chapter 2 Development of a viral method for Translating Ribosome Affinity Purification 39

2.1 Introduction 39

2.1.1 Transduction of neurons by viral vectors 39

2.1.2 Present work 43

2.2 Methods 45

2.2.1 Virus production 45

2.2.2 Primary culture and transduction by lentivirus 47

2.2.3 Viral injections 48

2.2.4 Translating Ribosome Affinity Purification (TRAP) 49

2.3 Results 51

2.3.1 Preparation of AAV and Lentiviruses for TRAP 51

2.3.2 Expression of lentiviruses in primary hippocampal culture 54

2.3.3 Optimisation of viral expression in-vivo 59

2.3.4 TRAP with virally transduced cells 64
2.4 Discussion 67
2.4.1 Virus development 67
2.4.2 Lentivirus expression in-vitro 67
2.4.3 Optimisation of viral expression in-vivo 68
2.4.4 Piloting of TRAP and dissection 70
2.4.5 Conclusions 71
Chapter 3 : A method of translatome profiling during associative recognition memory formation 72
3.1 Introduction 72
3.1.1 RNA sample preparation 72
3.1.2 Data processing 74
3.1.3 Present work 76
3.2 Methods 78
3.2.1 Subjects 78
3.2.2 Experimental Design 78
3.2.3 Viral injections 78
3.2.4 Paired viewing 79
3.2.5 Object in Place task 84
3.2.6 Translating Ribosome Affinity Purification 85
3.2.7 Amplification and sequencing 85
3.2.8 Bioinformatics 86
3.3 Results 90
3.3.1 Profiling associative recognition memory 90
3.3.2 Preparation of RNAseq libraries 92
3.3.3 Quality control of sequencing data 97
3.3.4 Library composition 107
3.3.5 Expression of hippocampal cell type markers 109
3.4 Discussion 116
3.4.1 Experiments to profile associative recognition memory 116
3.4.2 Quality control of TRAP RNA samples 117
3.4.3 Quality of sequencing data 118
3.4.4 Expression of hippocampal marker genes 118
3.4.5 Conclusions 119

3.5 Supplemental Figures 121

Chapter 4 Profiling the CA1 translatome during associative recognition memory formation 122

4.1 Introduction 122

4.1.1 Statistical modelling of gene expression data 122

4.1.2 Biological insight from RNAseq data 127

4.1.3 Present work 130

4.2 Methods 132

4.2.1 Detection of differentially expressed genes 132

4.2.2 Functional interpretation 133

4.3 Results 135

4.3.1 Optimisation of differential expression analysis 135

4.3.2 Immediate early gene expression 145

4.3.3 Exploratory analyses of genes that are differentially expressed in response to the paired viewing protocol 148

4.4 Discussion 164

4.4.1 Methods for differential expression analysis 164

4.4.2 Immediate early gene expression 165

4.4.3 Protein-protein interaction network of differentially expressed genes and putative hub genes 167

4.4.4 Clustering of co-expressed genes 168

4.4.5 Summary 171

4.5 Supplemental information 173

Chapter 5 General Discussion 177

5.1 TRAP viruses 177

5.2 Learning in the paired viewing paradigm 177

5.3 Differential expression following paired viewing 178

5.4 Limitations 179

5.5 Validation of findings and future work 180

5.6 Quantitative methods to discover molecular mechanisms underlying learning 181

5.7 Summary 182

References 183

Appendix 1: Solutions for TRAP 222

List of Figures

Figure 1.1: Diagram of hippocampal circuitry from Deng et al (2010). 3
Figure 1.2: Spontaneous exploration tests of recognition memory. 8
Figure 1.3: Induction of plasticity in response to stimulation. 17
Figure 1.4: Illumina Sequencing-by-synthesis method. 24
Figure 1.5: Methods of profiling the translatome (from King & Gerber, 2014). 27
Figure 2.1: Bands of DNA from restriction digests of viral plasmids 52
Figure 2.2: qPCR data for titre quantitation of lentiviruses CMV-EGFP-L10a and CaMKiiα-EGFP. 54
Figure 2.3: EGFP fluorescence in rat hippocampal cultures transduced by lentivirus.....55
Figure 2.4: Co-expression of cell type markers with lentivirus CMV-EGFP-L10a in hippocampal primary cultures.....57
Figure 2.5: Co-expression of cell type markers with lentivirus CaMKiiα-EGFP in hippocampal primary cultures.....58
Figure 2.6: Transduction of rat hippocampus by lentivirus and AAV expressing CMV-EGFP injected at varying co-ordinates. 60
Figure 2.7: Transduction of rat hippocampus by AAV and lentivirus expressing CMV-EGFP-L10a and lentivirus CaMKiiα-EGFP.....62
Figure 2.8: Transduction of CA1 by AAV and lentivirus CMV-EGFP-L10a; higher magnification and spread along AP axis.....63
Figure 2.9: Distribution of cell type markers in mouse hippocampus; in-situ hybridisation (ISH) images from Allen Mouse Brain Atlas.....64
Figure 2.10: HEK293T cultured cells transduced by lentivirus CMV-EGFP-L10a at 10x magnification.....64
Figure 3.1: Schematic of SeqPlex RNA Amplification kit procedure. 73
Figure 3.2: Schematic of the Trinity transcriptome assembly pipeline 76
Figure 3.3: Schematic of experimental design. 78
Figure 3.4: Schematic of paired viewing setup with an example pair of novel and familiar arrangements. 79
Figure 3.5: Labelled photographs of paired viewing setup. 80
Figure 3.6: Samples of stimuli used for paired viewing. 81
Figure 3.7: Schematic of novel, familiar and no-image control conditions of paired viewing. 83
Figure 3.8: Schematic of the object- in-place test. 84
Figure 3.9 : Process for assembly of contaminant sequences from raw sequencing data. 88
Figure 3.10: Fos expression following paired viewing of novel and familiar arrangements. 91
Figure 3.11: Mean discrimination ratio (DR) for exploration of items in novel vs familiar locations by rats injected with AAV CMV-EGFP-L10a into CA1. 92
Figure 3.12: Representative electrophoretic traces from purified TRAP-generated RNA samples indicating RNA integrity. 93
Figure 3.13: RNA concentration plotted against RNA integrity number equivalent (RIN^e) for all TRAP samples.....94

Figure 3.14: Consistency of amplification of selected genes by two commercially available amplification kits measured by qPCR. 96
Figure 3.15: Representative examples of FastQC graphs of sequence content. 98
Figure 3.16: Representative examples of FastQC graphs of sequence duplication and k-mer content. 99
Figure 3.17: Effects of RNA integrity number (RIN) and RNA concentration on alignment. 100
Figure 3.18: Comparison of Tophat2 and STAR splice-aware alignment tools.....101
Figure 3.19: Percentage of aligned reads in unbound RNA samples compared to immunoprecipitate samples. A greater proportion of reads were aligned and uniquely aligned in unbound (UB) samples compared to the corresponding immunoprecipitate (IP) samples. 101
Figure 3.20: Representative examples of FastQC graphs of sequence content after pre-processing.....104
Figure 3.21: Representative examples of FastQC graphs of sequence duplication and k-mer content after pre-processing.....105
Figure 3.22: Representative dupRadar plots of duplication rate against read density ..107
Figure 3.23: Alignment of reads to the rat genome and other sources. 108
Figure 3.24: Expression of neural cell type markers within TRAP samples. 111
Figure 3.25: Expression of markers of hippocampal subregions and CA1 domains in sequenced samples. 115
Figure 3.26: Expression of key neural cell type markers in individual samples.....121
Figure 3.27: Expression of selected markers of hippocampal subregions and CA1 domains in individual samples. 121
Figure 4.1: Distribution of relative $\log_2$ expression values (RLEs) before and after upper quartile (UQ) normalisation.....135
Figure 4.2: Principal components analysis (PCA) of PV-TRAP samples before and after RUVg normalisation. 136
Figure 4.3: Comparison of differential expression analyses.....138
Figure 4.4: Heatmap of differentially expressed genes (DEGs) identified by all tests across all conditions with hierarchical clustering. 140
Figure 4.5: Representative expression profiles, protein-protein interactions, and functions of the differentially expressed genes (DEGs) identified by all tests (N=54)....142
Figure 4.6: Correlated expression of postsynaptic density (PSD) proteins following paired viewing. 145
Figure 4.7: Expression of key immediate-early genes (IEGs) in PV-TRAP experiment. ...146
Figure 4.8: Differential expression of immediate early genes (IEGs) and delayed primary response genes. 147
Figure 4.9: Differentially expressed genes (DEGs) identified by edgeR robust likelihood ratio test across all conditions 149
Figure 4.10: Network of differentially expressed genes (DEGs) following paired viewing. 151
Figure 4.11: Putative hub genes within the network of differentially expressed genes (DEGs) following paired viewing. 153

Figure 4.12: K-means clustering of expression profiles of differentially expressed genes by stimulus condition. 157
Figure 4.13: Map of enriched gene ontology (GO) terms in clusters of genes with similar expression profiles for novel and familiar conditions. 161

List of Tables

Table 1.1: Comparison of lentivirus and AAV properties 33
Table 2.1: Viruses produced and viral backbone plasmids cloned in house 51
Table 2.2: Restriction digests to confirm cloning of viral backbone plasmids 51
Table 2.3: Sequencing results for cloned plasmids prior to virus production. 53
Table 2.4: Primary antibodies used to stain primary hippocampal cultures transduced by lentivirus. 56
Table 2.5: Co-ordinates of viral injections to optimise targeting of CA1 in rats. 59
Table 2.6: TRAP RNA from cultured HEK293T cells expressing EGFP-L10a 65
Table 2.7: TRAP RNA purified from dissected hippocampi of rats injected with AAV CMV-EGFP-L10a in pilot tests of TRAP method 66
Table 3.1: Paired viewing training procedure 81
Table 3.2: Counterbalancing of paired viewing conditions 82
Table 3.3: Quality control statistics for sequenced RNA samples used for analysis 103
Table 3.4: Overrepresented sequences and sources 106
Table 3.5: Cell type markers 110
Table 3.6: Markers of hippocampal subregions and CA1 domains 113
Table 4.1: Number of samples in each condition 132
Table 4.2: Number of differentially expressed genes detected using different methods and thresholds 139
Table 4.3: Top five enriched biological process GO terms among common differentially expressed genes 143
Table 4.4: Putative hub genes in the network of differentially expressed genes following paired viewing 155
Table 4.5: Contingency tables for genes with clustered expression profiles in novel, familiar, and control data 159
Table 4.6: Top five enriched GO terms in clusters of genes in the no-image control data 163
Table 4.7: Differentially expressed genes detected by all pipelines 173

Abbreviations

AAV; rAAV	Adeno-associated virus; recombinant AAV
ABC	Avidin-biotin complex
AMPAR	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor
AP	Anterior-posterior
Arc	Activity-regulated cytoskeletal protein
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BP	Biological process (GO term)
CA1; CA2; CA3	Cornu ammonis 1; 2; 3
CaM	Calmodulin
CaMK	Calcium/calmodulin-dependent kinase
cAMP	Cyclic adenosine monophosphate
CBA	Chicken beta-actin
CC	Cellular compartment (GO term)
Ccnh	Cyclin H
Cdk	Cyclin-dependent kinase
cDNA	complementary DNA
CFC	Contextual fear conditioning
CMV	Cytomegalovirus
CNS	Central nervous system
CpG	Cytosine-phosphate-guanine site
CPM	Counts per million
CREB	cAMP response element binding protein
Ct	Cycle threshold
DAB	3,3'-diaminobenzidine
DAPI	4’,6-diamidino-2-phenylindole
dCA1	dorsal CA1
ddH2O	Double-distilled water
DE	Differential expression
DEG	Differentially expressed gene
DG	Dentate gyrus
DHPC	Diheptanoyl-sn-glycero-3-phosphocholine
div	Days in vitro
dNTPs	Deoxynucleotides
DR	Discrimination ratio
dsDNA	double-stranded DNA
E18	Embryonic day 18
EC	Entorhinal cortex
eEF	Eukaryotic elongation factor
EF1a	Elongation factor 1a
eIF	Eukaryotic initiation factor
E-LTP	Early-phase long-term potentiation
FC	Fold change

FDR	False discovery rate
fMRI	Functional magnetic resonance imaging
FUV	Factor of unwanted variation
GABA	Gamma-aminobutyric acid
GAD	Glutamate decarboxylase
GC	Guanine-cytosine
GFAP	Glial fibrillary acidic protein
GI	Genome identifier
GLM	Generalised linear model
GO	Gene ontology
HIV	Human immunodeficiency virus
ICC	Immunocytochemistry
IHC	Immunohistochemistry
IN	Inhibitory interneuron
IP	Immunoprecipitate
ISH	In-situ hybridisation
ITR	Inverted terminal repeat
KO	Knock-out
lat.	Latitude
lincRNA	Long intergenic non-coding RNA
L-LTP	Late-phase long term potentiation
LRT; LRTr	Likelihood ratio test; LRT robust
LTD	Long-term depression
LTM	Long-term memory
LTP	Long-term potentiation
LTR	Long terminal repeat
MAPK	Mitogen-activated protein kinase
MeCP2	Methyl CpG binding protein 2
MF	Molecular function (GO term)
mGluR	Metabotropic glutamate receptor
MLV	Murine leukaemia virus
MWM	Morris water maze
mPFC	Medial prefrontal cortex
mRNA	Messenger RNA
miRNA	Micro-RNA
MTL	Medial temporal lobe
NB	Negative binomial
NCBI	National Center for Biotechnology Information
NGS	Next-generation sequencing
NMDAR	N-methyl D-aspartate receptor
NOR	Novel object recognition
Npas4	Neuronal PAS domain protein 4
NSE	Neuron-specific endolase
nt	Nucleotide
OiP	Object-in-place task

PB	Phosphate buffer
PBS; PBST	Phosphate buffered saline; PBS with Triton-X
PCA; PC	Principal components analysis; Principal component
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
Phka1	Phosphorylase kinase, alpha subunit
PN	Pyramidal neuron
Pol2	RNA polymerase II
PolyA	Poly-adenylated
PoRh	Postrhinal cortex
PP1; PP2A	Protein phosphatase 1; 2A
PPI	Protein-protein interaction
PRh	Perirhinal cortex
pRRL	A lentivirus transfer vector
PSD	Postsynaptic density
Psd95	Postsynaptic density protein 95
PSI	Protein synthesis inhibitor
PTM	Post-translational modification
PVIN	Parvalbumin interneuron
QC	Quality control
QL	Quasi-likelihood
qPCR	Quantitative polymerase chain reaction
Rfc4	Replication factor C 4
RIN; RIN^e	RNA integrity number; RIN equivalent
RLE	Relative log expression
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
ROI	Region of interest
RPKM	Reads Per Kilobase of transcript per Million mapped reads
rRNA	ribosomal RNA
RTF	Regulatory transcription factor
RUVseq	Remove unwanted variation from RNAseq data
SD	Standard deviation
SE	Single-end
SEM	Standard error of the mean
SLM	Stratum lacunosum moleculare
SO	Stratum oriens
SP	Stratum pyramidale
SR	Stratum radiatum
SRF	Serum response factor
ssDNA	Single-stranded DNA
SSE	Sum of squared errors
STM	Short-term memory
TF	Transcription factor
TRAP	Translating ribosome affinity purification

tRNA	Transfer RNA
TSS	Transcription start site
TSSM	Tromethamine, sodium chloride, sucrose and mannitol
tTa	Tetracycline-controlled transactivator
UTR	Untranslated region
UQ	Upper quartile
VSV-G	Vesicular stomatitis virus glycoprotein
WGCNA	Weighted gene co-expression network analysis
WPRE	Woodchuck post-transcriptional regulatory element
WTA	Whole transcriptome amplification

Chapter 1 Introduction

"Has it ever struck you ... that life is all memory, except for the one present moment that goes by you so quickly you hardly catch it going? It's really all memory... except for each passing moment."
-- Eric Kandel

Recognition is a form of memory that enables us to judge prior occurrence and thereby make sense of a complex world. Visual object recognition has been described as "perhaps the major adaptive function of vision, our dominant sense" (Cooke & Bear, 2015). Without the ability to determine that a stimulus has been encountered before, we would not be able to associate stimuli with their consequences, learn appropriate responses, identify changes in our environment, generalise to similar situations, or navigate.

Memories are encoded in the brain by modification of synaptic and neuronal properties (Martin et al, 2000; Whitlock et al, 2006; Wang & Morris, 2010) and long-term forms of memory and plasticity usually require synthesis of new proteins (Davis & Squire, 1984; Minatohara et al, 2016). The formation of associative recognition memories requires the hippocampus (Barker & Warburton, 2011; Warburton & Brown, 2015) and involves protein synthesis in hippocampal subregion CA1 (Wan et al, 1999; Aggleton et al, 2012). In the present work, a genome-wide translatome profiling study was conducted to investigate the identities, functions, and time-courses of genes differentially expressed in CA1 neurons within three hours of presentation of novel and familiar configurations of visual stimuli.

1.1 Memory and the brain

1.1.1 Types of memory and memory processes

Memory, the storage and recollection of past experience, is distributed across different regions of the brain and involves multiple separable processes. The hippocampus in the medial temporal lobe (MTL) is particularly important for memory formation and is crucial for the acquisition of explicit forms of memory including spatial and recognition memory (Cohen & Squire, 1980; Squire, 1992; Eichenbaum & Cohen, 2014). In humans, declarative memories are consciously accessible and include semantic memory for facts and episodic memory for events. Non-declarative forms of memory are independent of the hippocampus and include procedural motor memories and priming (Squire, 1992). This thesis will explore the intracellular events that accompany recognition memory formation.

The key stages of memory processing are encoding, consolidation, and recollection. Experimental and theoretical evidence suggests that memories are encoded in neuronal ensembles (Wilson & McNaughton, 1994; Blum & Abbott, 1996; Reijmers et al, 2007; Josselyn et al, 2015). Consolidation, the stabilisation of a memory trace and the formation of long-term memory (LTM, lasting longer than three hours; Gold & McGaugh, 1975; Izquierdo & Medina, 1997), occurs firstly at the level of the synapse within hours of acquisition of memory information and then subsequently at the systems level, as memories formed in the hippocampus are transferred to neocortex for long-term storage (McNaughton & Morris, 1987; Squire & Alvarez, 1995; Dudai & Morris, 2000). Prior to such systems-level consolidation, the reactivation of a memory trace during recollection has been shown to lead to the memory becoming labile and requiring reconsolidation to prevent memory loss (Nader et al, 2000; Milekic & Alberini, 2002; Barnes et al, 2012). Short-term memory (STM), in contrast to LTM, is considered to last between one and three hours and does not involve synaptic consolidation.

1.1.2 Neural pathways for memory and regional functions

1.1.2.1 Processing of visual stimuli

Visual information is derived from light detected by photoreceptors in the retina, from which signals are transferred to the brain via the optic nerve. In rodents, 97% of optic nerve fibres project to lateral geniculate nucleus in the contralateral hemisphere and from there are relayed to primary visual cortex (V1; Sefton et al, 2004), where neurons are specialised to respond to a range of specific properties including position, orientation, length, and movement of the object being viewed (Hubel & Wiesel, 1962; Poggio et al, 1985).

In both rodents and primates, information from V1 is hypothesised to be divided into two processing streams: the dorsal (or "where") stream, concerned with the perception of space, motion, and the control of actions, and the ventral (or "what") stream, primarily involved in object recognition and visual perception (Mishkin & Ungerleider, 1982; Goodale & Milner, 1992). The dorsal stream travels to V2, V3, and the middle temporal visual area, where cells are specialised for directional selectivity, and then to posterior parietal cortex. The ventral stream travels through visual areas V2 and V4 to the inferotemporal cortex, where neurons respond to complex forms.

Both streams of visual information provide input to the MTL memory system, which in rodents includes the perirhinal (PRh), postrhinal (PoRH), and entorhinal (EC) cortices and the hippocampus (Squire & Zola-Morgan, 1991; Burwell & Amaral, 1998). The dorsal visual stream preferentially projects to PoRH and medial EC, whereas the ventral stream preferentially projects to PRh and lateral EC (Kneirim et al, 2006). The EC has forward and backprojections to PRh and PoRH and provides the primary input to the hippocampus (Burwell & Amaral, 1998). These pathways thus convey the visual input necessary for the encoding of visual recognition memories.

1.1.2.2 Hippocampus

1.1.2.2.1 Anatomy

The hippocampus is divided into three major subfields, the anatomical connections of which form a tri-synaptic loop: dentate gyrus (DG), the CA3, and CA1. Figure 1.1 (from Deng et al, 2010) illustrates the layout of rodent hippocampus and major input and output pathways. DG and CA3 receive EC layer II input via the perforant pathway, whereas CA1 receives EC layer III input via the temporoammonic pathway. DG projects to CA3 via the mossy fibres and CA3 projects to CA1 via the Schaffer collaterals. CA1 is the main output structure of the hippocampus and projects primarily to the subiculum and EC, as well as to medial prefrontal cortex (mPFC). Thus the main flow of information processing through the hippocampus is unidirectional.

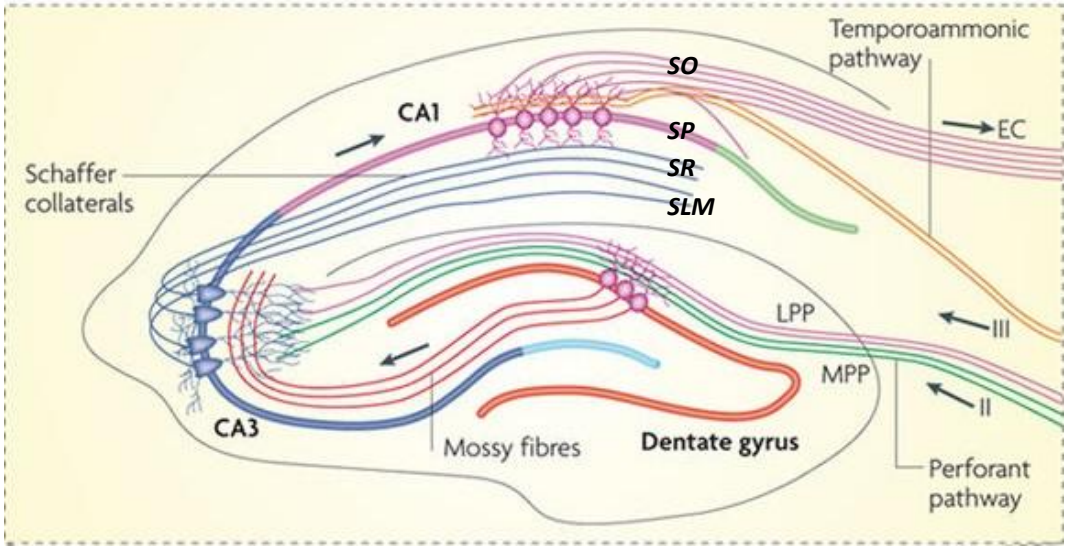


Figure 1.1: Diagram of hippocampal circuitry from Deng et al (2010). The hippocampal subfields dentate gyrus (DG), CA1 and CA3 are arranged in a trisynaptic loop. DG and CA3 receive input from entorhinal cortex (EC) layer II via the perforant pathway, which has both lateral and medial branches (LPP and MPP). DG cells project to CA3 via the mossy fibres. CA3 neurons have strong recurrent connections and project to CA1 via the Schaffer collaterals. CA1 additionally receives input from EC layer III via the temporoammonic pathway. CA1 then projects back to EC. Labels were added to indicate CA1 layers: stratum oriens (SO), stratum pyramidale (SP), stratum radiatum (SR), and stratum lacunosum-moleculare (SLM).

Including indirect projections via the subiculum and EC, the hippocampus is connected to many regions including the PFC, the thalamus and hypothalamus, midbrain regions such as the striatum and nucleus accumbens, and the amygdala (Bird & Burgess, 2008). Hippocampal activity is also regulated by neuromodulatory inputs received from different regions of the brain, including acetylcholine from the medial septum, noradrenaline from locus coereleus, dopamine from the midbrain, and serotonin from the Raphe nuclei in the brainstem (Rho & Storey, 2001). In light of the extensive connections of the hippocampus with a large number of brain regions, the hippocampus integrates multiple signals and propagates signals throughout the brain, facilitating the encoding and retrieval of activity patterns present during memory formation as well as the gradual consolidation of memories in neocortex.

Within CA1, excitatory pyramidal neurons (PNs) make up ~90% of neurons (Bezaire & Soltesz, 2013) and have their somas in the stratum pyramidale, one of four layers in CA1

(see Figure 1.1). Each PN has many basal dendrites in stratum oriens (SO) and one long branched apical dendrite, which extends into stratum radiatum (SR) and stratum lacunosum-moleculare (SLM). PNs in area CA3 project to both SR and SO, EC projects to SLM, and sparse local collaterals are received in SO (Freund & Buzsaki, 1996).

Inhibitory interneurons (INs) synapse with both PNs and other INs and help to maintain a balance in overall excitability and generate and maintain oscillatory activity (Klausberger & Somogyi, 2008; Ognjanovski et al, 2017). INs have diverse properties and can be classified based on their morphology, location, gene expression profiles, firing patterns, and connections with other neurons. More than twenty types of IN have been identified in CA1 (Freund & Buzsaki, 1996; Klausberger & Somogyi, 2008).

1.1.2.2.2 Hippocampal functions

The hippocampus is particularly important for the encoding and retrieval of spatial and contextual information. Memory tasks with a spatial or associative component activate dorsal CA1 (dCA1) and are impaired by lesions or inactivation of the hippocampus, including the Morris water maze (MWM; Morris et al, 1986; Riedel et al, 1999), contextual fear conditioning (CFC; Kim et al, 1992; Phillips & LeDoux, 1994), and the object-location and object-in-place recognition tasks (OiP; Bussey et al, 2000; Mumby et al, 2002; Barker & Warburton, 2011; see section 1.1.3.3).

The distinctive circuitry of the hippocampus and its subregions, discussed above, is suited for particular kinds of information processing. In highly influential computational research, David Marr (1971) proposed that the hippocampus performs pattern separation and pattern completion. According to this theory, the DG is specialised for pattern separation and integrates disparate inputs to generate sparse, orthogonal representations for processing by CA3. In CA3, modification of the dense recurrent connections between neurons enables different ensembles that are active during encoding to become associated, so that memories can later be retrieved by pattern completion from a fragment of the same activity (McNaughton & Morris, 1987; Rolls, 1996). CA1, which has many more neurons than CA3, may restore information removed during processing by associating co-occuring inputs, also enabling generalisation across similar representations (Treves & Rolls, 1994).

The hippocampus has also been shown to be critical for processing spatial information (Bird & Burgess, 2008) and contains a population of cells which have been shown to encode an animal's location within an environment, the place cell (O'Keefe & Dostrovsky, 1971; O'Keefe & Nadel, 1978). Indeed, ~20% of CA1 PNs have been shown to respond to spatial location and the place fields of individual cells have been shown to evolve over time (Ziv et al, 2013). Thus, over weeks of recording, 15-25% of place cells responded to the same location in consecutive sessions (Ziv et al, 2013). Place cell firing has also been shown to be orientation-independent and maintained in the dark (Quirk et al, 1990), so such firing is not straight-forwardly determined solely by visual input. Another type of neuron with location-dependent firing is the grid cell, found in the EC. Place cells can be contrasted with grid cells as the latter have hexagonal receptive fields that expand or shrink depending on the size of the current environment (Quirk et al,

1992), whereas place cells largely have a single place field. As an object is repeatedly experienced in a specific location, some of the "place cells" that initially responded to that location become responsive to the conjunction of the location and object (Moita et al, 2003; Komorowski et al, 2009).

In addition to spatial coding, cells in the CA1 also code for temporal information (Macdonald et al, 2011; Eichenbaum, 2014) and consistent with this firing is the finding that the hippocampus is crucial for the memory of sequences (Fortin et al, 2002; Kumaran & Maguire, 2006). Ensembles of "time cells" fire sequentially to code for successive moments and change their response properties when temporal aspects of a task, such as the length of a delay period, are updated (Macdonald et al, 2011). Many cells in the hippocampus respond differentially to both time and location, as well as other properties including object identity and behavioural responses (Gothard et al, 1996; Pastalkova et al, 2008; Macdonald et al, 2011).

An alternative perspective on CA1 function is the suggestion that this subfield of the hippocampus functions for match/mismatch detection by comparing internal hippocampal representations from CA3 with sensory information relayed from EC (Vinogradova & Dudaeva, 1972; Gothard et al, 1996; Kumaran & Maguire, 2006). CA1 PNs have increased firing rates during exposure to objects in novel locations (O'Keefe, 1976; Fyhn et al, 2002) and it has been suggested that this enables the broadcast of a generalised novelty signal from the CA1 to other brain regions (Larkin et al, 2014). This function may therefore be important for remembering/expecting sequences, recollecting related information, and mediating the preferential encoding of novel stimuli. It has been debated whether mismatch detection is performed by CA3 or CA1 (e.g. Lisman & Otmakhova, 2001; Vinogradova, 2001). For example, Vinogradova (2001) proposed that the CA3 acts as a comparator and Lee et al (2005) found that lesions of DG and CA3 but not CA1 severely impaired memory for object-location associations. Lee et al (2005) suggested that the CA3 performs mismatch detection for recently acquired memories, but that the CA1 may be required for comparison of present surroundings with older, consolidated memories.

To integrate different theories of hippocampal function, Cohen & Eichenbaum (1993) proposed that the hippocampus is specialised for relational processing and can form associations between stimuli of any modality (see also Eichenbaum & Cohen, 2014; Schiller et al, 2015). In this view, spatial information is encoded as a set of associations between locations (Milivojevic & Doeller, 2013) and analogous "cognitive maps" may be formed that represent abstract conceptual dimensions such as social hierarchies (Tolman, 1948; Tavares et al, 2015).

1.1.2.2.2.1 Divisions between dorsal and ventral hippocampus

In addition to the functional segregation within the hippocampal subfields, there is also strong evidence of functional divisions between dorsal and ventral hippocampus, with dorsal hippocampus more strongly involved in spatial processing and ventral hippocampus implicated in emotional processing including fear-related memory (Moser & Moser, 1998; Fanselow & Dong, 2010). In support of this distinction, there are more

place cells in dorsal than ventral hippocampus (Jung et al, 1994), navigational experience is correlated with dorsal but not ventral hippocampal volume in humans (Maguire et al, 2000), and memory acquisition in the MWM is abolished by dorsal hippocampal lesions but unaffected by lesions of ventral hippocampus (Moser et al, 1993). Association of a fear-inducing stimulus such as electric shock with a context is dependent on sufficient prior exploration (Fanselow & Dong, 2010) and is impaired by lesions of the dorsal hippocampus, but formation of tone-shock associations (cued fear) is not (Kim et al, 1992; Hunsaker & Kesner, 2008). Differences in gene expression along the dorsal-ventral axis have also been found in CA1 PNs (Dong et al, 2009). In this thesis the focus of experiments is on gene expression changes within the CA1 subregion of the dorsal hippocampus.

1.1.3 Recognition memory

1.1.3.1 Processes involved in recognition memory

Item recognition memory enables judgements of whether individual objects have been encountered before. Assessing the familiarity of stimuli is important as novel stimuli and events are likely to be behaviourally relevant and are preferentially encoded (Sokolov & Vinograda, 1975; Stark & Okado, 2003). However, items are rarely remembered in isolation, but can also be recognised in conjunction with other items, contexts or locations, by a process called associative recognition. While individual objects can be reliably identified following a single presentation (Standing, 1973), the learning of contextual associations requires multiple exposures (Smith & DeCoster, 2000; O'Reilly & Norman, 2002). In addition, it is possible to judge how recently stimuli were last encountered using recency memory (Brown & Aggleton, 2001; Warburton & Brown, 2015). Thus, recognition memory is not a unitary process.

Experimental evidence suggests that different forms of recognition are processed differently in the brain. For visual object recognition, form discrimination is required to determine object identity and a comparison with stored information enables a determination of familiarity. Associative recognition requires additional spatial and associative processing, whereas recency memory may require temporal processing. According to dual-process models, recognition memory involves two separable types of processing: familiarity discrimination and recollection (Mandler, 1980; Yonelinas, 1994). Familiarity discrimination occurs rapidly and involves the simple determination of whether a stimulus has been seen previously, whereas recollection is a slower process including the retrieval of additional contextual information (Brown & Aggleton, 2001). Object recognition can therefore be achieved using either familiarity or recollection-based processing, whereas associative recognition requires recollection.

1.1.3.2 Behavioural induction of recognition memory in rodents

Recognition memory can be investigated in rodents using spontaneous exploration tasks such as the widely used novel object recognition task (NOR; Ennaceur & Delacour, 1988). These tasks rely on the preference of rodents to explore novel over familiar stimuli, an innate behaviour that therefore does not require training (Ennaceur & Delacour, 1988; Mumby, 2001). The NOR task is outlined in Figure 1.2A: the subjects

explore two identical objects in an arena during the sample phase and after a delay period subjects are returned to the arena for the test phase in which one familiar object has been replaced by a novel object. Recognition memory in the rodents can be conveniently assayed by measuring differential exploration of novel and familiar stimuli in the test phase and calculating the discrimination ratio (DR; Albasser et al, 2010a). The effects of pharmacological, genetic, or surgical interventions on encoding, retrieval, and maintenance of recognition memory can be investigated by measuring changes in the DR and the delay period can be varied to study STM and LTM as required.

Other spontaneous exploration tasks include the object-location, object-in-place (OiP), and temporal order tasks (Dix & Aggleton, 1999; Warburton & Brown, 2015), shown in Figure 1.2B-D respectively. In the object-location task, exploration of familiar objects is compared in novel versus familiar locations. In this task, differential exploration may be driven by the novelty of the location or by the novelty of the object-location association. In the OiP task, four objects are placed in four corners of an arena and the positions of two objects are swapped during the delay period. Rodents with intact OiP memory will spend more time exploring the objects in new positions. The OiP task probes associative recognition memory as all the individual objects and locations in the test phase are familiar, but associations with proximal and distal location cues differ. The temporal order task probes recency memory by presentation of two pairs of identical stimuli separated by a delay in the sample phase, followed by presentation of one stimulus from each pair. Rodents with recency memory intact will on average spend more time exploring the object that was presented first. The advantage of these tasks is that different forms of recognition memory may be measured using the same apparatus and behavioural indices.

Although these tasks have been used in a number of studies to investigate the neural bases of recognition memory (e.g. Winters et al, 2004; Barker & Warburton, 2011), they cannot be applied to the scientific questions that are the focus of this thesis. It is difficult to detect subtle differences in RNA and protein levels between conditions using spontaneous exploration paradigms due to high variability between subjects within the same condition, affected by factors including differing levels of motor activity, motivation, alertness, and concentrations of circulating hormones. Additionally, multiple presentations of object stimuli may be required to induce robust changes in a sufficiently large number of cells, yet only a single stimulus or arrangement can be tested within a session in the arena-based NOR and OiP tasks without removing the animal from the apparatus, which is time consuming and may increase stress levels (Aggleton et al, 2012). Hence the bow-tie maze was developed by Aggleton and colleagues to assess recognition memory by presenting multiple objects using a shuttle box in which one end of the arena can be blocked off while rearrangements of the objects are made for additional trials (Albasser et al, 2010a). However, there are also limitations in using the bow-tie maze to assess gene expression due to variability between subjects, hence in the present studies the paired viewing (PV) procedure was used to present series of novel and familiar images to rodents, as outlined in the following section.

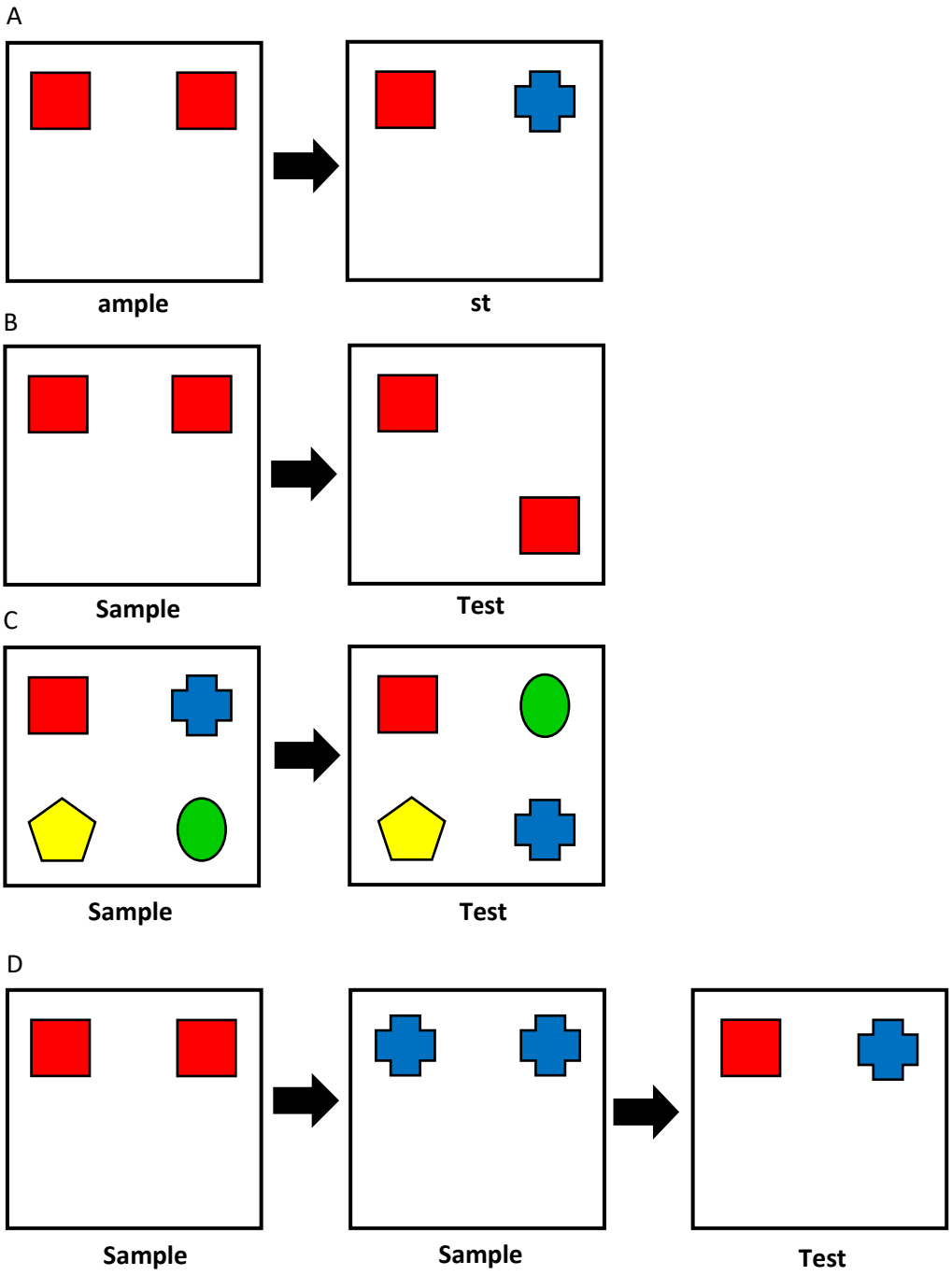


Figure 1.2: Spontaneous exploration tests of recognition memory. A: Novel object recognition. B: Object-location. C: Object-in-place. D: Temporal order

1.1.3.2.1 Paired viewing

The paired viewing procedure was developed by Zhu et al (1996) to compare neural activation in response to novel and familiar conditions within individual subjects, thus it is possible using this procedure to control for most sources of unwanted variability, such as might occur when using between-subjects protocols. This control is achieved by training an animal to stay in place during image presentation by delivery of juice reward. Novel images (which may be items or arrangements) are presented to one eye while familiar images are simultaneously presented to the other eye using two screens

separated by a divider to occlude the binocular field. This arrangement for presentation of the stimuli is possible because rats and mice have laterally facing eyes and a largely monocular field of view (Wallace et al, 2013).

In rodents, retinal ganglion cells project almost exclusively to the contralateral hemisphere (Jeffrey, 1984; Sefton et al, 2004), such that significant differences in the response to stimuli can be observed between hemispheres depending on visual input (Montero & Jian, 1995; Wan et al, 1999; Kemp & Manahan-Vaughan, 2012). As the comparison of neural activity between novel and familiar stimuli is made within the same animal, recognition is passive and cannot be behaviourally measured. Nevertheless, differences between novel and familiar conditions in gene expression and plasticity processes have been observed using immunohistochemistry (IHC; Wan et al, 1999) and electrophysiological recordings (Kemp & Manahan-Vaughan, 2012) respectively. The lack of direct interaction with stimuli may enable a purer study of visual recognition memory compared to spontaneous exploration tasks as comparison of paired images does not incorporate multisensory processing, motor behaviour, or decision-making which may affect memory and plasticity processes (Winters & Reid 2010; Aggleton et al, 2012).

1.1.3.3 Pathways and regions involved in recognition memory

Research into the neural basis of recognition memory using the behavioural paradigms discussed above has implicated a circuit consisting of PRh, hippocampus, mPFC, and more recently, mediodorsal thalamus in the acquisition and retrieval of item and associative recognition memory (Aggleton et al, 2012; Kinnavane et al, 2015; Warburton & Brown, 2015).

Convergent evidence from rodent lesion studies (Mumby & Pinel, 1994; Langston & Wood, 2010; Barker & Warburton, 2011), pharmacological interventions (Barker et al, 2006; Barker & Warburton, 2008; Brown et al, 2012), electrophysiology (Zhu et al, 1995; Brown & Aggleton, 2001) and immediate early gene imaging (Zhu et al, 1996; Wan et al, 1999; Aggleton et al, 2012) suggests a double dissociation between neural regions involved in the processing of item recognition and processing of associations and spatial information necessary for associative recognition. Thus, lesions of the PRh have been shown to impair NOR and OiP memory but not object-location memory (Winters et al, 2004; Barker et al, 2007; Jo & Lee, 2010; Brown et al, 2012), likely due to PRh involvement in processing object identity and object familiarity (Warburton & Brown, 2015), while lesions of the hippocampus have been shown to disrupt object location and OiP memory but not NOR (Bussey et al, 2000; Mumby et al, 2002; Barker & Warburton, 2011).

Further evidence of the role of different brain regions in different recognition memory processes has been provided by immunohistochemical studies. Expression of Fos protein as a marker of neuronal activation (see section 1.2.3.1) has been measured 90 minutes after stimulus presentation in the PV task (Zhu et al, 1996; Wan et al, 1999). Significantly more Fos-stained nuclei were found in PRh and area TE following viewing of novel versus familiar items, while more Fos-stained nuclei were found in CA1 and PoRH

following viewing of novel versus familiar arrangements (Wan et al, 1999). Significant differences following paired viewing of arrangements were also found in the dentate gyrus (DG) and subiculum, but interestingly Fos levels were higher in the familiar condition in these regions. Similarly, in humans viewing spatial rearrangements of stimuli has been shown to increase activity measured by functional magnetic resonance imaging (fMRI) in dorsal hippocampus (Pihlajamäki et al, 2004) and lesions of the macaque dorsal hippocampus have been found to impair preferential viewing of novel spatial configurations of stimuli but not of novel items or items in novel locations in a visual paired comparison task (Bachevalier & Nemanic, 2008).

Conflicting results have been found regarding whether the hippocampus plays a role in item recognition and the topic is heavily disputed (Clark et al, 2000; Barker & Warburton, 2011; Warburton & Brown, 2015; Cohen & Stackman, 2015). A recent meta-analysis (Cohen & Stackman, 2015) found that NOR was not impaired in most lesion studies (e.g. Ainge et al, 2006; Barker & Warburton, 2011), whereas most temporary inactivation studies have found impairment in NOR after delay periods ≥10 minutes (e.g. Baker & Kim, 2002; Cohen et al, 2013). One interpretation of these findings is that the hippocampus is involved in the storage of NOR memory under normal conditions, but other regions can compensate for this function in the event of long-term damage to the hippocampus (Cohen & Stackman, 2015). Nevertheless, the hippocampus has been shown to be activated during object exploration, potentially due to involvement of multisensory stimulus processing (Ainge et al, 2006; Aggleton et al, 2012). Following the bowtie maze task, Fos expression was found to be increased in CA1 and CA3 in the novel condition, but behavioural performance was not impaired by hippocampal lesions (Albasser et al, 2010a) and is impaired by PRh lesions (Horne et al, 2010). This finding suggests that presentation and exploration of novel objects in different locations in the bow-tie maze results in the activation of hippocampal neurons, but the behavioural output is driven by item recognition.

The hippocampus is also involved in recency memory (Barker & Warburton, 2011; Albasser et al, 2012). More specifically, CA1 lesions but not CA3 lesions have been found to impair performance in the temporal order task (Hoge and Kesner, 2007) and a recent study found that inactivation of projections from dorsal CA1 to mPFC, but not from intermediate CA1 to mPFC, impaired temporal order memory (Barker et al, 2017). Lesions of mPFC and the mediodorsal thalamus also impair recency and OiP memory but not NOR memory (Barker et al, 2007; Cross et al, 2013), although more work is needed to understand the roles of these regions in associative recognition (Warburton & Brown, 2015).

In humans, most research supports a dual-process model for recognition memory in which familiarity is more associated with information processing in the PRh, whereas recollection is more associated with hippocampal function (Brown & Aggleton, 2001). For example, some hippocampal lesion patients have deficits in associative forms of recognition memory but spared item recognition memory (Vargha-Khadem et al, 1997; Mayes et al, 1999). An alternative approach that has been applied to distinguish between familiarity discrimination and recollection in the study of recognition memory

is signal detection theory, using receiver-operating characteristics (ROCs; Yonelinas, 1994). ROCs can be used to compare the proportions of correct and incorrect recognitions in a recognition task at varying levels of confidence. Yonelinas (1994) proposed that symmetrical ROCs were consistent with familiarity discrimination, whereas ROCs that were skewed towards a higher proportion of correct recognitions at high confidence levels must involve a different process, i.e. recollection. In rats, evidence of both familiarity and recollection components of ROCs was found and the recollection component was abolished by lesioning of the hippocampus (Fortin et al, 2004).

Cellular firing properties inform the functions of different regions. In PRh and IT cortex, three populations of neurons have been described, differentiated by their changes in firing rates to novel and familiar stimuli, the so called 'novelty', 'familiarity', and 'recency' neurons (Xiang & Brown, 1998; Brown & Aggleton, 2001). Novelty neurons have been shown to respond maximally to the first encounter of an object, whereas familiarity neurons have reduced responses to highly familiar stimuli but do not respond differently to the first and second presentations. Recency neurons have reduced responses to stimuli that have been seen recently regardless of overall familiarity. Hippocampal neurons, as discussed in section 1.1.2.2.2, respond to a range of properties including spatial location (O'Keefe & Nadel, 1978), temporal information (Macdonald et al, 2011), and conjunctions of properties including object identity, reward associations, and behavioural responses in addition to place and time (Pastalkova et al, 2008; Macdonald et al, 2011; McKenzie et al, 2014). Some hippocampal neurons show increased firing in response to changes in the spatial locations of objects (Rolls et al, 1989; Lenck-Santini et al, 2005; McKenzie et al, 2014). The rearrangement of stimuli activates overlapping ensembles, whereas exposure to different environments activates largely separate ensembles (Vazdarjanova & Guzowski, 2004). Only a small proportion of hippocampal cells have reduced responses to repeated object-location combinations (Rolls et al, 1989; Brown & Aggleton, 2001), but novelty shifts the theta-phase of PN firing in CA1 (Jeewajee et al, 2008; Lever et al, 2010) and this mechanism may support enhanced synaptic potentiation for encoding of recognition-related information (Hölscher et al, 1997; Lever et al, 2010).

Interactions between PRh, hippocampus, and mPFC may be necessary to support specific forms of recognition memory (Warburton & Brown, 2015). For example, during the OiP task, inactivation of PRh disrupts firing patterns in CA1 (Lee & Park, 2013). Hippocampal activity is also strongly correlated with mPFC activity during the OiP task (Kim et al, 2011). Changes in functional connectivity may differentially support different kinds of processing. Application of structural equation modelling to patterns of Fos expression during the bow-tie maze has shown that EC input to DG is favoured during exploration of novel stimuli, whereas EC input to CA1 is favoured during exploration of familiar stimuli (Albasser et al, 2010b; Aggleton et al, 2012). In humans, an fMRI study found that PRh activity precedes hippocampal activity during familiarity-based recognition, whereas PRh activity is modulated by the hippocampus during recollection (Rugg & Vilberg, 2013). The strength of connectivity between the two regions was positively correlated with performance using both processes. Thus, although the

hippocampus is crucial to the acquisition of associative recognition memories, these results suggest that memory storage may be distributed across the circuit and not confined to any one region (Warburton & Brown, 2015).

1.2 Genes and cellular processes in the control of synaptic plasticity and memory formation

Synaptic plasticity, the modification of connection strength between neurons, is widely thought to be the primary mechanism that underlies learning and memory (Morris & Frey, 1997; Martin et al, 2000; Whitlock et al, 2006; Kandel et al, 2014). Hebbian plasticity, proposed by Donald Hebb (1949), describes the strengthening of connections between cells with coincident firing such that subsequent stimulation of the presynaptic neuron will be more likely to excite the postsynaptic neuron. Long-term potentiation (LTP) of synaptic strength was first demonstrated experimentally by Bliss & Lomo (1973) in the hippocampus and has been recorded in behaving animals during learning (Whitlock et al, 2006). Synaptic plasticity is now known to have multiple mechanisms, including long-term depression (LTD; Dudek & Bear, 1992; Malenka & Bear, 2004) and homeostatic mechanisms (Abraham & Bear, 1996; Citri & Malenka, 2007; Lisman, 2017).

Changes in gene expression have been found to be necessary for most forms of enduring plasticity, including late-phase LTP (L-LTP; Davis & Squire, 1984; Alberini, 2009; Okuno, 2011; Minatohara et al, 2016). De novo protein synthesis occurs following exposure to novel stimuli and during learning (Davis & Squire, 1984; Guzowski et al, 1999; 2001) and LTM formation is impaired by the administration of protein synthesis inhibitors (PSIs) immediately before a training session and approximately three hours later, while memory acquisition and STM formation are spared (Barondes & Cohen, 1967; Greksch & Matthies, 1980). LTM has similar sensitive periods to impairment by inhibitors of glycoprotein synthesis (Tiunova et al, 1998), protein kinase A (PKA) activity leading to transcription initiation (Bourtchuladze et al, 1998), and RNA Polymerase II, an enzyme required for transcription (Igaz et al, 2002). Administration of PSIs twelve hours after inhibitory avoidance training has been shown to impair memory retention seven days but not two days later, suggesting that protein synthesis at later timepoints is required for persistent LTM (Bekinschtein et al, 2007). Reconsolidation is also blocked by PSI administration (Nader et al, 2000; Debiec et al, 2002; Rossato et al, 2007). Studies of associative recognition memory have shown that long-term object-in-context memory is impaired by infusion of PSIs into the hippocampus but not perirhinal cortex (Balderas et al, 2008), and that differential gene expression is induced in CA1 following viewing or exploration of objects in novel locations (Wan et al, 1999; Albasser et al, 2010a).

Plasticity processes are regulated at multiple levels: from extracellular signalling, to receptor activation, intracellular protein-protein interactions, altered gene expression, and protein and mRNA trafficking and stability (Malenka & Bear, 2004; Bolognani & Perrone-Bizzozero, 2008; Miyashita et al, 2008; Fioravante & Byrne, 2011). Different forms of plasticity can be induced at the same synapses by different patterns of stimulation (Markram et al, 1997; Bi & Poo, 1998; Mizuno et al, 2001; Edelmann et al, 2017) and involve overlapping mechanisms (Mayford et al, 1995; Wang et al, 2005; Yger & Gilson, 2015). The same type of plasticity (e.g. LTP) can also often be induced by more than one mechanism within the same cells (Minichiello et al, 2002; Malenka & Bear, 2004). Different plasticity processes have been found to contribute to different types

and stages of memory (Kemp & Manahan-Vaughan, 2004; Mizuno & Giese, 2005; Miyashita et al, 2008; Minatohara et al, 2016; Edelmann et al, 2017) and so behavioural studies are important to determine the mechanisms involved in particular types of learning. The following sections outline the main forms of plasticity in CA1 neurons, the activity-dependent regulation of gene expression, and the role of differential gene expression in plasticity processes.

1.2.1 Types of plasticity

1.2.1.1 Hebbian plasticity

Hebbian plasticity is the potentiation or depression of the number and strength of synaptic connections and affects subsequent synaptic transmission. Fast synaptic transmission is principally mediated by α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors (AMPARs), ionotropic glutamate receptors that are present in dendritic spines. Hebbian plasticity can be mediated by changes at both presynaptic and postsynaptic sites (Sutton & Carew, 2000; Palmer et al, 2004; Kandel et al, 2014): at presynaptic terminals, plasticity is expressed by modification of neurotransmitter release and calcium channel properties (Mellor et al, 2002; Catterall & Few, 2008), while at postsynaptic sites, plasticity is expressed through changes to the number, conductance, and calcium permeability of AMPARs (Beattie et al, 2000; Bredt & Nicoll, 2003; Malenka & Bear, 2004) and the number and properties of dendritic spines (Yuste & Bonhoeffer, 2001; Kasai et al, 2003; Matsuzaki et al, 2004; Bosch & Hayashi, 2012; Giese et al, 2015).

LTP and LTD are induced in response to different stimuli (Kemp & Manahan-Vaughan, 2004; Kandel et al, 2014). In slice electrophysiology experiments, stimulation at 100Hz in CA1 induces LTP, whereas prolonged low-frequency stimulation and high-frequency stimulation after induction of LTP both induce LTD (Dudek & Bear, 1992; Barr et al, 1995; Beattie et al, 2000; Kandel et al, 2014). At lower stimulation frequencies closer to physiological levels, whether LTP or LTD is induced depends on the order of presynaptic and postsynaptic firing and this is referred to as spike-timing dependent plasticity (Markram et al, 1997; Bi & Poo, 1998; Nevian & Sakmann, 2006). At the molecular level, differential induction of LTP and LTD is mediated by Ca^{2+} influx and receptor activation (Mizuno et al, 2001). In behavioural experiments, LTP and LTD have been found to mediate different types of learning. In the amygdala for example, learned fear has been found to require synaptic potentiation (Rumpel et al, 2005), whereas learned safety has been found to require LTD (Rogan et al, 2005). Kemp & Manahan-Vaughan (2004) found that exploration of a novel environment without objects facilitated LTP in CA1, whereas exploration of an environment containing either novel objects or familiar objects in novel locations facilitated LTD, suggesting that LTD may underlie object-in-place memory (Manahan-Vaughan and Braunewell, 1999) and both forms of plasticity may be required to encode complete memories integrating different types of information (Kemp & Manahan-Vaughan, 2007).

The most studied form of plasticity at CA3-CA1 synapses is LTP dependent on the activity of N-methyl D-aspartate receptors (NMDARs; Bliss & Collingridge, 1993). NMDARs are

activated by coincident firing of presynaptic and postsynaptic neurons, enabling associations to be formed between neurons that are active at the same time, because depolarisation is required to remove a Mg^{2+} ion that blocks the NMDAR channel. NMDARs are also required for a form of LTD (Dudek & Bear, 1992; Oliet et al, 1997) and there are forms of both LTP and LTD that do not require NMDAR activation but depend on metabotropic glutamate receptor (mGluR) activation (Bashir et al, 1993; Malenka & Bear, 2004; Wang et al, 2016). Metabotropic receptors enable response amplification as substrate binding activates a G-protein, which initiates signalling cascades leading to downstream opening of disparate ion channels.

NMDAR channels are permeable to Ca^{2+}, an essential second-messenger that is involved in the initiation of multiple intracellular signalling pathways, including activation of Ca^{2+}/calmodulin-dependent protein kinase 2 (CaMKii; Lisman et al, 2002). One function of CaMKii is the phosphorylation of AMPAR subunits, thereby increasing the number of functional AMPARs in the postsynaptic membrane and generating E-LTP (Barria et al, 1997; Hayashi et al, 2000; Lee et al, 2000). The activation of CaMKii can be sustained by autophosphorylation and this mechanism is required for memory and LTP (Giese et al, 1998; Lisman et al, 2002). Additional AMPARs are also inserted into the membrane during L-LTP, supported by increased recycling of early endosomes (Park et al., 2004) and causing enlargement of spine heads (Matsuzaki et al, 2004). Conversely, synaptic depression is mediated by dephosphorylation (Lee et al, 1998; 2000) and clathrin-mediated endocytosis of AMPARs (Lüscher et al, 1999; Beattie et al, 2000). Synaptic strength is also modified by structural plasticity processes, including remodelling of the actin cytoskeleton to alter spine morphology and number (Yuste & Bonhoeffer, 2001; Matsuzaki et al, 2004; Bosch & Hayashi, 2012; Caroni et al, 2012) and adhesion to strengthen contacts between pre- and postsynaptic neurons (Benson et al, 2000).

Experimental evidence has shown that both LTP and memory formation have an initial protein synthesis-independent phase, early-phase LTP (E-LTP), that is mediated by protein-protein interactions and a later phase, L-LTP, that requires de novo gene expression and is induced independently (Matthies et al, 1990; Frey et al, 1993). An intermediate phase of LTP that is dependent on local protein synthesis but does not require transcription has also been identified (Ghirardi et al, 1995; Sutton & Carew, 2000; Raymond et al, 2000). Experimental evidence suggests that Ca^{2+} signals from different sources and associated with different cellular compartments may be required for different phases of LTP (Raymond & Redman, 2002; Raymond, 2007). For example, while NMDARs are implicated in all three phases of LTP (Raymond, 2007), the activation of Group I mGluRs has been found to initiate translation-dependent (intermediate) LTP by triggering the release of internal Ca^{2+} stores in distal dendrites (Nakamura et al, 2000). Local protein synthesis, but not de novo transcription, is also required for mGluR-dependent LTD but not for NMDAR-dependent LTD (Huber et al, 2000; Snyder et al, 2001).

1.2.1.2 Homeostatic and other forms of plasticity

Computational modelling studies have consistently found that networks implementing Hebbian plasticity mechanisms require additional homeostatic forms of plasticity to

maintain both network stability and the potential for further plasticity (e.g. Rochester et al, 1956; Marder & Prinz, 2002; Renart et al, 2003; Zenke et al, 2013). Homeostatic plasticity has also been demonstrated experimentally in response to changes in synaptic activity (Turrigiano et al, 1998; Burrone et al, 2002). For example, neuronal activity has been shown to be upregulated when visual input is reduced (Kaneko et al, 2008). Homeostatic plasticity involves both synapse-specific and intrinsic mechanisms that regulate firing rate, subthreshold activity, and synaptic weight (Mozzachiodi & Byrne, 2010; Keck et al, 2017) and may be involved in memory formation at multiple timescales (Zenke & Gerstner, 2017).

Two proposed forms of homeostatic plasticity are synaptic scaling, the regulation of absolute synaptic weights while relative weights are maintained (Turrigiano, 2007), and metaplasticity, the regulation of thresholds for plasticity induction (Abraham, 2008). Like Hebbian forms of plasticity, homeostatic scaling can involve a range of presynaptic and postsynaptic mechanisms, including regulation of AMPAR trafficking and local synthesis of new AMPAR proteins (Ju et al, 2004; Pozo & Goda, 2010). Astrocytes have also been found to regulate homeostatic scaling by release of tumour necrosis factor alpha (TNFα; Stellwagen & Malenka, 2006). Metaplasticity can be mediated by changes to the number, activity and subtypes of NMDARs (Macdonald et al, 2006) and regulated at both synaptic and global levels (Abraham, 2008). De novo transcription and translation have been shown to be required for some forms of homeostatic plasticity (Davis & Goodman, 1998; Ibata et al, 2008; Lyons & West, 2011).

Taken together, the findings discussed in this section illustrate the diversity of plasticity processes and the dependence of some forms of plasticity on the regulation of gene expression.

1.2.2 Regulation of gene expression for plasticity

The induction of gene expression for enduring forms of plasticity requires the initiation of intracellular signalling cascades by second-messengers induced by receptor activity including Ca^{2+} and cyclic adenosine monoamine triphosphate (cAMP; Lynch et al, 1983; Malenka et al, 1988; Salin et al, 1996). Signalling pathways regulate gene expression at the local level by translation of existing mRNA populations at synapses (Steward & Levy, 1982; Kang & Schuman, 1996; Sutton & Schuman, 2006) and at the global level by mechanisms including the activation of activity-dependent regulatory transcription factors (RTFs; Sheng et al, 1991; Bito et al, 1996; Nonaka et al, 2014). RTFs induce the transcription of immediate early genes, which in turn regulate downstream gene expression and cellular processes (Sheng & Greenberg, 1990). Pathways from stimulation to plasticity are outlined in Figure 1.3, illustrating the roles of transcriptional and translational regulation in the expression of different forms of plasticity, as discussed in the previous section (Matthies et al, 1990; Huber et al, 2000; Bradshaw et al, 2003; Raymond, 2007). Regulatory processes are separated by timescale: in mammalian cells, post-translational modification (e.g. RTF activation) takes <1 second, transcription of a 10kbp gene takes ~10 minutes, translation of a protein with 300 amino acids takes ~1 minute, and protein transport is much slower, in the order of hours (Shamir et al, 2016).

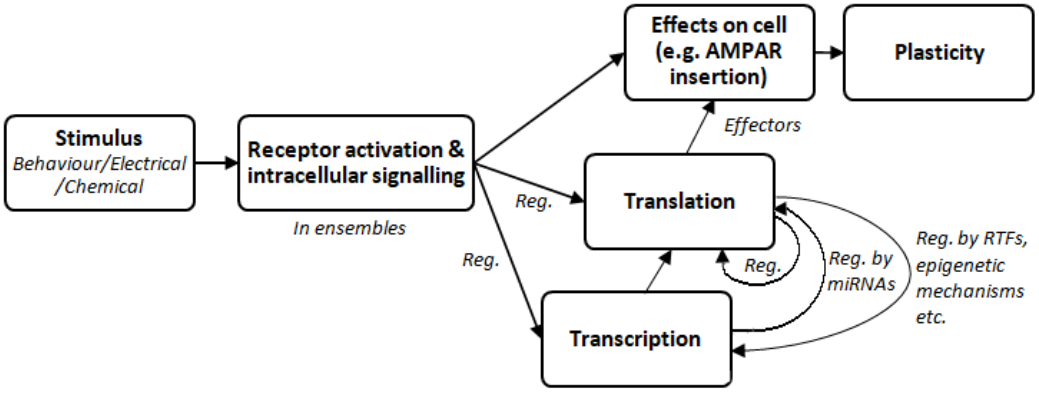


Figure 1.3: Induction of plasticity in response to stimulation. Different types of stimuli activate different receptors in neuronal membranes, inducing intracellular signalling pathways. Different signalling pathways induce different forms of plasticity by direct effects on the cell and by regulating the synthesis of new proteins, which can further affect cellular and synaptic properties. New proteins are synthesised both from existing populations of mRNAs and de-novo transcribed mRNAs. Transcription and translation are further regulated (reg.) by newly synthesised proteins and miRNAs. For example, transcription can be regulated by regulatory transcription factors (RTFs) and by epigenetic mechanisms, such as DNA methylation. The mechanisms regulating translation and transcription are discussed further in sections 1.2.2.1 and 1.2.2.2.

The intracellular signalling cascades involved in synaptic plasticity comprise mainly kinases and phosphatases, which regulate enzyme activity by altering the phosphorylation status of proteins. Kinases, including CaMKii, mitogen-activated protein kinases (MAPKs), protein kinase A (PKA) and protein kinase C (PKC), catalyse the donation of a phosphate group from ATP to a substrate protein and are predominantly involved in LTP, whereas phosphatases catalyse the reverse process and are mainly involved in LTD (Mulkey et al, 1993; Izquierdo & Medina, 1997; Malenka & Bear, 2004; Racaniello et al, 2010). Different kinases are activated downstream of different receptors and regulate different downstream processes (Smolen et al, 2006; Coba et al, 2008; Kotaleski & Blackwell, 2010). The processes regulating transcription and translation are discussed in further detail below.

1.2.2.1 Regulation of translation

Ribosomes, which mediate translation, have two subunits (the 60S and the larger 80S) and are made up of four types of rRNA and 79 proteins. The first step in translation is the formation of the 43S ribosomal pre-initiation complex from the 40S ribosome, transfer RNA (tRNA), guanosine triphosphate (GTP), and eukaryotic initiation factors (eIFs; Costa-Mattioli et al, 2009). Messenger RNA is then recruited to the ribosome by the interaction of eIFs with structures at the mRNA 5' end (Shatkin, 1985). The ribosome traverses the mRNA to find the initiation codon (AUG) and joins with the 60S ribosomal subunit, again mediated by eIFs (Costa-Mattioli et al, 2009). Elongation of the polypeptide is regulated by eukaryotic elongation factors (eEFs).

Translation has been found to be controlled during learning by bidirectional regulation of the activities of eIFs and eEFs (Redpath et al, 1993; Gingras et al, 1999; Dever, 2002; Costa-Mattioli et al, 2007; Richter & Klann, 2009). Translation can also be inhibited by binding of micro-RNAs (miRNAs) to complementary sequences in mRNAs (Filipowicz et al, 2008). The expression of miRNAs has been found to be affected by synaptic activity and regulate memory and plasticity processes (Vo et al, 2005; Schratt et al, 2006; Scott

et al, 2012; Peixoto et al, 2015). Cis-acting elements within mRNAs also regulate translation, such as internal ribosome entry sites in the 5' untranslated region (UTR) that provide an alternative method of mRNA recruitment to the ribosome (Doudna & Sarnow, 2007) and upstream opening reading frames that cause early termination of translation (Gaba et al, 2001; Eacker et al, preprint). Translation is regulated primarily at the initiation stage by the targeting of mRNA subpopulations by cis-acting elements and PTM of eIFs (Costa-Mattioli et al, 2009; Schwanhäusser et al, 2011).

1.2.2.2 Regulation of transcription

Transcription is a global process initiated by the PTM of activity-dependent RTFs including cAMP response element binding factor (CREB) and serum response factor (SRF; Dash et al, 1990; Miranti et al, 1995; Nonaka et al, 2014). Different types of stimulation induce distinct gene expression programs (Bartel et al, 1989; Dijkmans et al, 2009; Lyons & West, 2011), mediated by induction of different signalling cascades (Hardingham et al, 2001; Flavell et al, 2008). For example, genes transcribed in hippocampal cultures in response to GABA-A receptor antagonism have been found to be enriched for genes known to be regulated by SRF (Iacono et al, 2013) and CREB signalling can be regulated by nuclear calcium released from intracellular stores alone (Hardingham et al, 2001). A modelling study found that interactions between the key kinase pathways may have combinatorial effects on mRNA transcription (Jain & Bhalla, 2014).

TFs are defined by the presence of DNA binding domains which recognise 6-12bp degenerate sequences (response elements) in the promoters of expressed genes. The core promoter contains binding sites for general TFs (e.g. TATA-binding protein (TBP) and TFIIB) and the transcription start site (TSS), where RNA polymerase II (Pol2) binds to initiate transcription of protein-coding genes. Unlike general TFs, RTFs bind at proximal promoters ~250bps upstream of the TSS, at distal promoters located further upstream, or at enhancers or silencers. These latter additional regulatory elements can be located several kilobases away from the genes they affect and require activators and repressors, respectively, to bring them into proximity with promoters to influence gene expression. Combinatorial regulation of gene expression results from binding of TFs at multiple promoter and enhancer sequences (Joo et al, 2016).

Activity-dependent regulation of transcription requires signals to be received by the nucleus, which may be relayed by signalling molecules such as MAPK that have been shown to translocate from the synapse to the nucleus following synaptic activity (Martin et al, 1997). However, this mechanism is too slow to underlie transcription of IEGs that occurs within minutes of stimulation (Guzowski et al, 1999; Saha & Dudek, 2013) and signals to the nucleus can instead be relayed by calcium induced by action potentials without synaptic involvement (Hardingham et al, 2001; Adams & Dudek, 2005). Initiation of transcription requires the binding of TFs at gene promoters to modify chromatin topology, assemble the pre-initiation complex, and recruit Pol2 to the TSS (Saha & Dudek, 2013). These steps take approximately ten minutes (Serizawa et al, 1997), but like translation, transcription can also be regulated at the elongation stage (Saha et al, 2011). Rapid transcription of IEGs has been found to be facilitated by promoter-proximal stalling of Pol2 (Saha et al, 2011) and induced by double-strand DNA breaks in IEG

promoters in response to synaptic activity (Suberbielle et al, 2013; Madabhushi et al, 2015). The consolidation of LTM also involves regulation of transcription by epigenetic mechanisms that modify chromatin topology, such as DNA methylation and PTM of histones (Guan et al, 2002; Levenson et al, 2004; Miller et al, 2008).

Some newly synthesised mRNAs are transported to stimulated synapses prior to translation by the interaction of motor proteins with the cytoskeleton (Steward & Worley, 2001; Bramham & Wells, 2007; Martin & Ephrussi, 2009). According to the synaptic tagging and capture hypothesis, the mRNAs of plasticity-related proteins transcribed in response to strong stimulation are captured by a 'tag' at activated synapses, enabling L-LTP to also occur at synapses that received concurrent weak stimulation (Frey & Morris, 1997; Redondo & Morris, 2011). Synaptic tagging has been found to involve many different molecules and processes and may require local protein synthesis (Richter & Sonenberg, 2005; Redondo & Morris, 2011).

The activity-dependent translatome thus includes newly-transcribed genes undergoing translation both in the soma and at specific synapses, as well as existing mRNAs regulated at the translational level.

1.2.3 Differential gene expression and plasticity

The first wave of genes to be transcribed and translated in the cellular response to external stimuli are the immediate early genes (IEGs), which are rapidly induced and do not require de-novo protein synthesis for their transcription (Sheng & Greenberg, 1990). An additional group of activity-regulated genes that are transcribed in the presence of PSIs have slower kinetics and are sometimes referred to as delayed primary response genes (Tullai et al, 2007). The expression of IEGs and other primary response genes can be regulated by differential activation of kinase cascades for the expression of different forms of plasticity (Coba et al, 2008). Secondary response genes, in contrast, require de-novo protein synthesis for their expression (Tullai et al, 2007).

1.2.3.1 Immediate Early Genes

An estimated 30-40 IEGs are involved in the neuronal response to stimuli in the hippocampus, of which the majority are RTFs that co-ordinate downstream gene expression and between ten and fifteen are effectors with direct involvement in cellular functions (Lanahan & Worley, 1998). RTF IEGs, which include Fos, Zif268 (Egr1), Nr4a1 (Nur77), and Npas4, drive expression of delayed effector genes and regulate the expression of other IEGs. For example, Npas4 and Fos bind to each other's enhancers (Kim et al, 2010; Malik et al, 2014). Effector IEGs, such as Arc, Homer1a, and Cox2, have a range of functions and include enzymes, structural proteins, growth factors, and proteins involved in signal transduction (Lanahan & Worley, 1998). Arc for example is a cytoskeletal protein that regulates trafficking of AMPARs by endocytosis (Chowdhury et al, 2006) and may be necessary for the stabilisation of different forms of structural synaptic plasticity (Waung et al, 2008). Homer1a is an intracellular signalling protein that is localised to excitatory synapses in hippocampal neurons and binds to group I mGluRs in response to activity (Xiao et al, 1998). It has been proposed that Arc and Homer1a both mediate homeostatic scaling by different mechanisms (Shepherd et al, 2006; Hu et

al, 2010b), in addition to roles in other forms of plasticity. Most effector IEGs, except for Arc, are induced slowly compared to RTF IEGs (Miyashita et al, 2008; Saha et al, 2011).

1.2.3.1.1 Roles of IEG expression in memory and plasticity

IEGs are expressed in a subset of neurons following the induction of plasticity (Cole et al, 1989) and exposure to novel stimuli (Guzowski et al, 1999; Ramírez-Amaya et al, 2005; Vazdarjanova et al, 2006), without requirement of interstimulus contingencies or reinforcements. Thus, IEGs such as Fos and Arc are often used as markers of neuronal activation to identify brain regions and/or neuronal ensembles involved in memory tasks (Guzowski et al, 1999; Tagawa et al, 2005; Okuno, 2011; Barry et al, 2016). Fos is a particularly useful marker as it has low basal expression, is strongly induced within 15 mins of stimulation, and is then degraded (Morgan et al, 1987; Aggleton et al, 2012). Fos mRNA expression peaks 30 minutes after stimulation and protein expression peaks after ~90 minutes (Morgan et al, 1987).

Inhibition of expression of the IEGs Fos and Arc by the infusion of antisense mRNA has been shown to impair performance in a variety of memory tasks (Guzowski et al 2001; Fleischmann et al, 2003; McIntyre et al, 2005; Plath et al, 2006). Furthermore, Arc inhibition has been found to disrupt the maintenance but not induction of LTP (Guzowski et al, 2001) and the maintenance of LTD (Plath et al, 2006), while CNS-specific knock-out (KO) of Fos has been found to impair LTP induction at CA3-CA1 synapses (Fleischmann et al, 2003). Expression of IEGs in the hippocampus has been shown to habituate over repeated exposures to stimuli (Nikolaev et al, 1992; Anokhin & Rose, 1990; Bertaina-Anglade et al, 2000) and correlate with variables related to memory, including performance on a water maze task (Guzowski et al. 2001), exploration in a novel place task (Mendez et al, 2015), and visual experience (Tagawa et al, 2005). IEG expression may also be required at later timepoints for memory consolidation as inhibition of Fos expression in dCA1 twelve hours after inhibitory avoidance training has been shown to impair retrieval seven days later (Katche et al, 2010). Recent experiments using optogenetics to manipulate neuronal ensembles expressing Fos or Arc following a behavioural task have shown that reactivation leads to memory retrieval and inactivation leads to memory suppression (Liu et al, 2012b; Tanaka et al, 2014), strengthening the hypothesis that ensembles of IEG-expressing neurons represent engrams (Josselyn et al, 2015).

Although the set of IEGs induced by different types of stimuli is broadly similar across cell types (Sun & Lin, 2016), different stages and types of memory and plasticity have been found to induce different sets of IEGs (Guzowski et al, 2001; Miyashita et al, 2008; Ramamoorthi et al, 2011; Heroux et al, 2018) and more IEGs are induced by more intense stimulation (Worley et al, 1993). Different IEGs have also been found to have different expression profiles (Bartel et al, 1989; Guzowski et al, 2001; Zangenehpour & Chaudhuri, 2002; Cheval et al, 2012). For example, a CFC study found that Arc and Fos were induced in CA3 in a shock-only condition, whereas Npas4 was expressed only in conditions that involved context learning (Ramamoorthi et al, 2011). Unlike Fos and Arc, Npas4 is expressed in both excitatory and inhibitory neurons and has been found to be involved in upregulation of GABAergic synapses onto excitatory neurons and L-LTP of

excitatory input to GABAergic neurons (Lin et al, 2008; Sun & Lin, 2016). Using a spatial water maze task, Guzowski et al (2001) found that Fos, Zif268, and Arc were all upregulated in the hippocampal transcriptome 30 minutes after training, but each showed different patterns of expression between day 1 and day 7 of training and after reversal learning on day 7: Zif268 expression declined the most from day 1 to day 7, whereas Arc was most strongly upregulated by reversal learning. However, few studies to date have systematically investigated the expression profiles of different IEGs (Minatohara et al, 2016) and further research is needed to understand the contributions of IEGs to distinct forms of plasticity.

1.2.3.2 Genome-wide expression profiling

The population of mRNAs in hippocampal cells has been profiled at different timepoints following memory tasks in genome-wide expression studies, revealing bidirectional regulation of the expression of large numbers of genes with a variety of functions, including regulators of gene expression (transcription, translation and chromatin structure), genes associated with the actin cytoskeleton and structural plasticity, metabolic genes, and intracellular and extracellular signalling molecules (Donahue et al, 2002; Cavallaro et al, 2002; Barnes et al, 2012; Iacono et al, 2013; Chen et al, 2017). Gene expression has been found to differ across types and stages of memory (Miyashita et al, 2008; Poplawski et al, 2014) and expression profiling studies have suggested that a subset of the genes involved in consolidation are induced during reconsolidation (von Hertzen & Giese, 2005; Barnes et al, 2012). In a microarray study comparing gene expression in the hippocampus following acquisition, retrieval, and learning a new platform location (extinction/new learning) in a spatial water maze task, some genes (including many IEGs) were found to be regulated in the same direction in all groups after 30 minutes compared to home cage controls, whereas other genes were regulated only in some conditions (Miyashita et al, 2008). Based on these findings, the authors proposed that “core” plasticity genes may be activated at all stages of learning, whereas other genes involved in plasticity processes may be “state specific”. Gene expression has also been found to be regulated in response to other elements of behavioural tasks, including water deprivation (Almague-Melian et al, 2003), stress (Cullinan et al, 1996; Barry et al, 2016), and motor activity (Cavallaro et al, 2002). Matched control conditions are therefore important to determine whether IEG expression is due to learning or to other aspects of the task (Cavallaro et al, 2002; Shires & Aggleton, 2008).

As illustrated by the research discussed above, memory-specific studies are necessary to understand the plasticity processes and pathways involved in different forms of learning. In the present study, differential gene expression in the translatome of CA1 neurons was profiled to determine the key genes, including IEGs, expressed during associative recognition memory formation. The expression of IEGs was also used as a positive control to ensure that changes in gene expression in response to the experimental conditions were observed. The timepoints profiled were 10 minutes after learning, corresponding to the initial induction of differential gene expression; 30 and 90 minutes, corresponding to the peak of Fos mRNA and protein expression respectively

(Morgan et al, 1987), and 180 minutes, corresponding to the induction of a second wave of protein synthesis (Greksch & Matthies, 1980).

1.3 Profiling genome-wide gene expression

1.3.1 Identifying mRNA transcripts

High-throughput methods of identifying mRNA transcripts enable simultaneous profiling of all genes expressed in population of cells. These methods can be more informative than studying proteins in isolation as proteins have meaningful functions only within a network of other proteins and molecules (D'haeseleer et al, 2000). Protein-protein interaction networks have been shown to contain high levels of redundancy and feedback to ensure that the cell is robust to minor fluctuations and perturbations (Ma'ayan et al 2005). High-throughput methods are therefore an invaluable tool for the investigation of cellular processes.

The main two methods of genome-wide mRNA profiling are microarrays and RNA sequencing (RNAseq). Microarrays provide a measure of the relative quantities of known genes in a sample. Several thousand genes can be profiled in parallel using probes fixed to an array that bind to complementary sequences in mRNAs of target genes. RNAseq decodes the sequence of nucleotides in RNA fragments and so this method does not require prior knowledge of which sequences may be present and enables detection of sequence variation such as alternative splicing events and single nucleotide polymorphisms, as well as of non-coding sequences such as miRNAs. As sequencing costs have lowered, RNAseq has largely superseded microarrays (van Dijk et al, 2014). In the present study, RNAseq was conducted using the leading next-generation sequencing (NGS) platform, Illumina (CA, USA).

1.3.1.1 Next-generation sequencing

NGS methods were developed in 2005 following the sequencing of the human genome and enabled faster, cheaper, high-throughput sequencing of millions of short reads in parallel as opposed to individual fragments of DNA (van Dijk et al, 2014). The Illumina platform uses the sequencing-by-synthesis method described below (Bentley et al, 2008).

1.3.1.1.1 Library preparation

Prior to sequencing, purified RNA is first reverse transcribed to produce more stable complementary DNA (cDNA) and cDNA is fragmented using one of several methods including enzyme digestion, physical methods such as acoustic shearing, and chemical methods. For Illumina single-end (SE) sequencing, fragments of ~200 base pairs (bp) are recommended. The ends of fragments are blunted or repaired, a phosphate group is attached at the 5' end, and a poly-adenine (poly-A) tail is ligated or extended at the 3' end to promote cDNA stability and to facilitate the ligation of adapters for sequencing (Head et al, 2014). Following adapter ligation, cDNA libraries are amplified using the adapter sequences as primers for polymerase chain reaction (PCR) to enrich for product with ligated adapters. Adapters may contain a short barcode sequence so that multiple samples can be sequenced on the same flow cell. Excess primers can be removed by PCR purification and larger fragments such as adapter dimers can be removed by size selection using gel electrophoresis or paramagnetic beads.

1.3.1.1.2 Sequencing-by-synthesis method

In the first stage of the sequencing-by-synthesis method (Bentley et al, 2008), adapters bind to complementary oligonucleotides fixed to a flow cell. Each captured fragment undergoes bridge amplification as illustrated in Figure 1.4A: adapters at the unbound end of fragments bind to nearby complementary oligonucleotides and a parallel strand is synthesised. This process is repeated to form clonal clusters of identical fragments as shown in Figure 1.4B(ii). Deoxynucleotides (dNTPs) conjugated to fluorophores are then incorporated and visualised one at a time, as shown in Figure 1.4B(i,iii). The flow cell is imaged and base identity at each clonal cluster is determined by the colour or wavelength of light emitted. The number of cycles then determines the length of the read and the number of clusters determines the number of reads. Sequences and per-base quality scores are decoded by a base-calling algorithm.

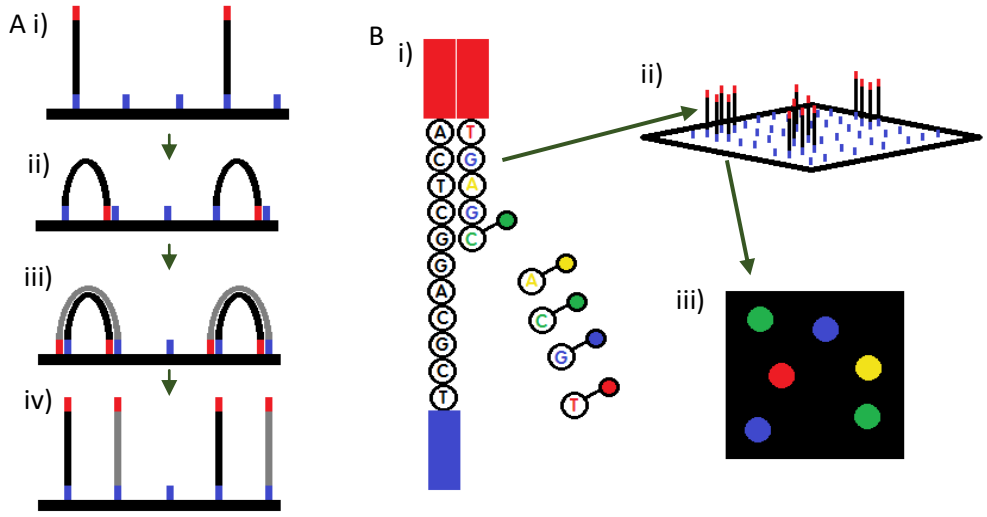


Figure 1.4: Illumina Sequencing-by-synthesis method. A: Bridge amplification. i) An adapter at one end of a library fragment binds to a complementary oligonucleotide on a flow cell; ii) the unbound adapter binds to a nearby complementary oligonucleotide; iii) a parallel strand is synthesised, and iv) strands are separated. B: Sequence detection. i) Deoxynucleotides conjugated to fluorophores are incorporated and visualised one at a time. ii) Each fragment belongs to a clonal cluster of identical fragments formed by repeated bridge amplification. iii) The flow cell is visualised and base identity can be determined by the colour of light emitted by a clonal cluster (e.g. in this example green corresponds to cytosine (C)).

1.3.1.1.3 Data processing and analysis

Raw data produced by NGS pipelines is in the form of millions of reads, 50-250 nucleotides in length, with associated quality scores. Quality scores for Illumina sequencing are assigned using the Phred+33 system in which numbers between 1 and ~40 correspond to the log probability that a given base call is incorrect, determined from peak resolution and shape. For example, a score of 20 indicates an error probability of one in one hundred.

After quality control and removal of low-quality reads and bases, reads are interpreted by alignment to a genome or assembly into consensus sequences. Alignment of RNAseq data must account for gaps in relation to the reference genome where introns have been removed by splicing during the formation of mature mRNA. Splice-aware aligners

include TopHat2 (Kim et al, 2013) and STAR (Dobin et al, 2013). Alternatively, where there is no reference genome that can be used for alignment, transcripts can be assembled directly from the data (Grabherr et al, 2011).

Aligned reads mapping to features can then be quantified using a quantification tool such as Cufflinks (Trapnell et al, 2010) or HTSeq (Anders et al, 2015), prior to downstream analysis of differential expression between samples. Analysis of alternative splicing (the expression of different isoforms of the same gene) can also be conducted with sufficient depth of sequencing (Trapnell et al, 2010). A huge range of analyses are possible using genome-wide expression data. The primary goals of many genome-wide profiling experiments are to determine the identities of genes that are differentially expressed between conditions and the enrichment of genes or transcripts involved in particular molecular functions or pathways. Many tools are available for each stage of RNAseq data analysis using differing approaches (Conesa et al, 2016) and methods of analysis are undergoing continuous innovation (see further discussion of tools in sections 3.1.2, 4.1.2, and 4.1.1).

1.3.2 Translatome profiling

Profiling the translatome provides a more accurate estimate of protein abundance than whole-transcriptome profiling (as in RNAseq) as mRNA transcript abundance accounts for only an estimated 25-30% of variation (Vogel et al, 2010). This is because not all mRNA is translated due to regulatory processes occurring between transcription and translation (Goldie & Cairns, 2012; King & Gerber, 2014) and the transcriptome has low temporal specificity, particularly in rat brain where mRNA has an exceptionally long average half-life of 12.5 days (Bondy, 1966; von Hungen et al, 1968). In contrast, mRNAs are bound to ribosomes for only minutes during translation (Shamir et al, 2016). RNAseq samples also include pre-mRNA which has not been fully spliced or may have been incorrectly spliced and is therefore likely to undergo nonsense-mediated decay instead of translation (Lewis et al, 2003). Differences in alternative splicing in RNAseq samples may consequently be spurious, reflecting transcription errors and background noise, whereas differentially spliced transcripts in the translatome are more likely to correspond to functional proteins (Zhang et al, 2015). Furthermore, only a small proportion of the total mRNA population are undergoing translation at any given time and consequently translatome profiling has increased power to detect differential expression (Tebaldi et al, 2012; Chen et al, 2017). The CA1 neuron translatome was therefore targeted in the present genome-wide expression profiling study.

These findings suggest that translatome profiling is a more sensitive method than total RNAseq for the detection of differential expression between behavioural conditions and during learning.

1.3.2.1 Methods for translatome profiling

Methods of profiling the translatome depend on isolation of ribosome-bound mRNA. The first method of translatome profiling to be developed was sucrose density-gradient fractionation, which separates mRNAs bound to polysomes (multiple ribosomes) from ribosomal subunits and free RNA by molecular weight (Johannes et al, 1999; King &

Gerber, 2014). Fractionation can also be used to profile ribosome occupancy as polysomes are separated along a gradient by the number of bound ribosomes and this may reflect translational status (Arava et al, 2003).

More recently, methods of polysome profiling based on affinity purification have been developed, including translating ribosome affinity purification (TRAP; Heiman et al, 2008; Doyle et al, 2008) and RiboTag (Sanz et al, 2009). These methods have the major advantage of cell-type specificity and are also quicker, easier to implement, require a lower quantity of input RNA, and have lower levels of contamination by other molecules compared to methods using fractionation (King & Gerber, 2014). Alternatively, positions of individual ribosomes on translating mRNAs can be profiled to study translational regulation in greater detail (Ingolia et al, 2009), a method known as ribosome footprinting. The fractionation, ribosome footprinting, and affinity purification methods of translatome profiling are outlined the schematic in Figure 1.5 (from King & Gerber, 2014).

1.3.2.2 Translating ribosome affinity purification

TRAP (Heiman et al, 2008; Doyle et al, 2008; Heiman et al, 2014) was developed to profile polysome-bound mRNA in spatially overlapping cell populations by expression of a genetically defined ribosomal tag. In this method, a cell-type-specific promoter is used to drive expression of a transgene encoding a ribosomal protein (usually L10a from the large ribosomal subunit) fused to enhanced green fluorescent protein (EGFP). Tagged ribosomes are then immunoprecipitated using magnetic beads coated in antibodies to EGFP so that mRNAs undergoing translation can be selectively purified and identified using NGS or microarrays. Samples of input and unbound RNA can be purified alongside the immunoprecipitate (IP) to distinguish between genes that are differentially expressed in IP samples due to differences in transcription and genes that are differentially expressed due to post-transcriptional regulation (Doyle et al, 2008).

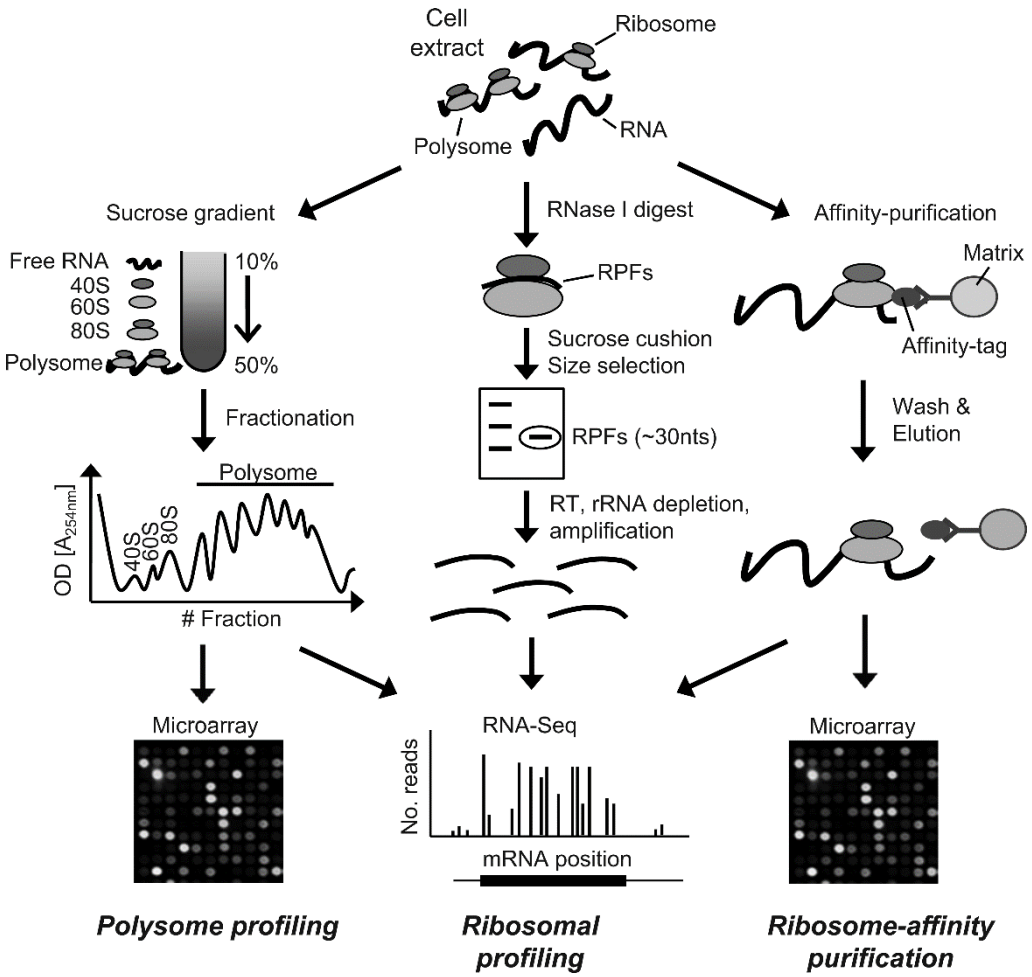


Figure 1.5: Methods of profiling the translatome (from King & Gerber, 2014). Left: mRNAs bound to polysomes can be separated from free RNA and ribosomal constituents by sucrose density-gradient fractionation. Centre: Ribosome protected fragments (RPFs) can be profiled by digestion of samples with RNase I, followed by size selection of fragments on a sucrose cushion and the reverse transcription and amplification of fragments. Right: mRNAs bound to polysomes can be purified by immunoprecipitation of ribosomes using an affinity tag and subsequently eluted. RNA can be profiled using microarrays or RNAseq, whereas profiling of RPFs requires RNAseq.

The TRAP method was developed using bacterial artificial chromosomes to generate transgenic mouse lines, but the EGFP-L10a transgene can also be delivered to neurons by direct injection of a virus (e.g. Kratz et al, 2014). Transgenic methods enable a range of complex expression patterns to target specific cell populations. For example, cell populations can be targeted by their projections using a retrogradely transported virus (Ekstrand et al, 2014; Cook-Snyder et al, 2015) or gene expression can be switched on and off using a tetracycline transactivator (tTa) system to express EGFP-L10a for limited periods of time in neurons activated by a behavioural paradigm (Drane et al, 2014). The tTa system could further be combined with a Fos promoter to profile gene expression in behaviourally-defined cell populations (Drane et al, 2014). This method is unfortunately not suitable to study the early stages of memory consolidation due to the length of time needed for EGFP-L10a to express and be incorporated into a functional ribosome, but the method could be used to study gene expression in neuronal populations activated by a behavioural paradigm at later timepoints such as during retrieval. An alternative method of affinity purification to target cells based on their activity takes advantage of

the phosphorylation of ribosomal protein S6 in response neuronal activation (Knight et al, 2012). In this protocol, pS6 is directly immunoprecipitated from dissected tissue to profile ribosome-bound mRNAs in activated cells, thereby enabling the identification of activated cell types or of cell-type markers so that those cells can be targeted in future studies (Knight et al, 2012). However, this method has a low yield compared to TRAP as the EGFP tag is more clearly defined and so is robustly immunoprecipitated. A drawback of TRAP is that not all ribosomes are tagged as expression of the transgene must compete with the native ribosomal protein (King & Gerber, 2014). Additionally, larger numbers of cells are required than for whole-transcriptome profiling as the RNA yield is relatively low (Jung & Jung, 2016).

1.3.2.3 Ribosome footprinting

Ribosome footprinting, also known as ribosome profiling, is a method to determine the positions of bound ribosomes on mRNAs and thereby investigate translational regulation by profiling short ribosome-protected fragments (~30 nucleotides) of mRNA (Ingolia et al, 2009). Samples for ribosome footprinting can be derived using either fractionation (Ingolia et al, 2009) or affinity purification (Ingolia et al, 2014; Eacker et al, preprint) and RNA not protected by ribosomes is degraded by treatment with RNaseI. Total RNA samples are sequenced in parallel so that fragments can be mapped to the transcriptome (Ingolia et al, 2009).

Analysis of ribosome footprinting data can reveal a wealth of information, including rates of mRNA translation (Tuller et al, 2011), novel initiation sites (Ingolia et al, 2009) and pause sites (Ingolia et al, 2011), and novel translated sequences (Calviello et al, 2016). Ribosome profiling has also provided evidence that long intergenic non-coding RNAs (lincRNAs) are bound by ribosomes and translated (Ingolia et al, 2011; 2014), a finding that has since been validated using other methods (Carlevaro-Fita et al, 2016). It has been proposed that lincRNA translation may provide a mechanism to regulate the lincRNA population (Carlevaro-Fita et al, 2016) and that the identities of bound lincRNAs may have functional relevance (Deniz & Erman, 2017). Furthermore, a small proportion of mRNA is bound by ribosomes but unlikely to be translated due to binding at upstream open reading frames leading to premature termination (Meijer & Thomas, 2002) and these mRNAs can be identified by ribosome footprinting (Ingolia et al, 2009).

One downside of ribosome footprinting is that data is difficult to map unambiguously to the genome due to the short length of fragments (King & Gerber, 2014). Artefacts in the positions of bound ribosomes can also be introduced by the protocol. For example, treatment with cycloheximide to stall translation may induce accumulation of ribosomes at initiation sites and therefore give misleading results (Gerashchenko & Gladyshev, 2014). Nevertheless, ribosome footprinting can provide much greater insight into the mechanisms of translational regulation, where this is the aim of an experiment, compared to methods of polysome profiling.

1.3.2.4 Present work

TRAP was used in this thesis to profile changes in the translatome of CA1 neurons as this method enables cell-type selectivity, is easy to apply compared to alternative methods,

and is easier to interpret than ribosome profiling. Differential mRNA-ribosome association has been shown to be detectable in CA1 PN dendrites within 10 minutes of a behavioural experience (Ainsley et al, 2014) and previous studies have demonstrated that TRAP is an effective method for time-course profiling (Huang et al, 2013), such as a recent study of gene expression in PNs in CA3-CA1 mini-slices that found greater differential gene expression between timepoints in TRAP samples compared to RNAseq samples (Chen et al, 2017).

1.4 Modification of gene expression in-vivo

To investigate the functions of genes in living cells, it is often desirable to modify gene expression by introducing transgenes. Transgenes are cloned into vectors for gene transfer into a target organism and gene expression is mediated within cells either by integration of the foreign DNA into host chromosomes or by DNA episomes outside the nucleus such as plasmids (closed circles of DNA). Genes can be transferred into eukaryotic cells by transfection with naked plasmid DNA or transduction by viral vectors. Transfection requires methods such as microinjection, electroporation, or heat-shock for DNA to cross the cell membrane, whereas viruses have evolved to infect host cells with much greater efficiency. Two major approaches to the modification of gene expression in laboratory animals will be discussed in this section: generating transgenic organisms and using viral vectors to deliver transgenes to individual subjects.

1.4.1 Transgenic organisms

Transgenic lines are produced by the introduction of genes into the germline, thereby modifying gene expression in an organism and its descendants. The two primary methods of generating transgenic animals are microinjection of DNA directly into the pronucleus of a fertilised egg and the transduction of embryonic stem cells in culture. In both methods, an embryo is formed and implanted into the uterus of a mouse and only a small proportion of offspring will express the transgene. Subjects that are heterozygous for the transgene can be mated to breed homozygotes if required. Highly specific expression can be driven in transgenic animals because stem cells can be transduced by bacterial artificial chromosomes (BACs), which unlike viral vectors have the capacity to carry hundreds of kilobases of DNA and therefore much larger genes and more extensive regulatory sequences (Heintz, 2001).

A transgenic mouse line has been bred to induce strong, specific expression of the TRAP gene in CA1 pyramidal neurons, as required in the present study, using a CaMKiiα promoter under control of a tTa system (Drane et al, 2014). However, impaired spatial reference memory and neurodegeneration have been reported in one of the lines cross-bred to produce these mice, the CaMKiiα-tTa line (Han et al, 2012). Moreover, in paired viewing pilot experiments, mice required multiple pre-training sessions as well as juice pre-exposure to learn to make head entries, did not reliably make the required number of head entries per session during training, and often exited the observation window after triggering a trial and so were not exposed to the full set of images. Rats, in contrast, reliably made the required number of head entries per session from the first exposure to the paired viewing apparatus and so were used as subjects in the present study.

The first transgenic rats were produced almost thirty years ago (Hammer et al, 1990; Mullins et al, 1990) and transgenic rats with modified gene expression in the hippocampus have been used to study learning and memory (Thorsell et al, 2000; Wang et al, 2009). However, it is much more difficult to generate transgenic rats than transgenic mice and until recently there have been few lines available. This is due to difficulties both with the survival of zygotes following micro-injection (van den Brandt,

2004) and the maintenance of rat stem cell cultures (Hamra et al, 2002). Viral vectors were therefore developed in the present study to express the EGFP-L10a transgene in rat hippocampus.

1.4.2 Viruses

Due to their ability to invade foreign cells and use the cell's machinery to express their own genes, viruses are an invaluable tool for the transduction of neurons in vivo. Viruses are formed of genetic material (DNA or RNA), enough to code for 4-200 proteins (Lodish et al, 2000), inside a protein coat called a capsid. The capsid consists of many copies of one or a small number of proteins and may be helical or icosahedral (20-sided). Covering the capsid, some viruses have a phospholipid envelope formed from a combination of the membrane of a host cell and viral glycoproteins when a new virion buds from the host cell. Different types of virus can infect animal, bacterial or plant cells, but most do not cross phyla and some only infect closely related species.

Viral vectors for gene transfer are produced by removing non-essential and potentially harmful sequences and cloning in a target gene. By using different types of viruses, viral vectors with different properties can be produced. Animal viruses can be classified according to the form of their genetic material (e.g. DNA or RNA, double or single stranded) as this is key to their life cycle (Lodish et al, 2000). Viruses have two types of life cycle: lytic and lysogenic. During a lytic cycle, viral particles are assembled within a host cell and then released to infect new cells. During a lysogenic cycle, viral genes are integrated into the host chromosomes and may remain dormant.

The first viral vectors were developed in the early 1980s and were simple retroviruses (Shimotohno & Temin, 1981; Bouard et al, 2009). Retroviruses are enclosed in a lipid membrane and contain two copies of their genome in the form of positive strands of RNA. Simple retroviruses infect dividing cells, where their RNA is reverse transcribed into DNA and incorporated into one of the host's chromosomes during mitosis. Viral genes are then expressed and replicated by the cell's own machinery in subsequent cell divisions. Retroviruses can therefore mediate stable lifelong expression in dividing cell populations. Some viral vectors can use the host ribosomes to express episomally and so can transduce both dividing and non-dividing cells. However, episomal expression in dividing cells is usually short term.

Viral vectors have undergone rapid development since their inception, with several types of virus commonly in use which are also modified and combined to improve their suitability for different applications. A list of seven desired properties of viral vectors for gene transfer is provided by Howarth et al (2010): ability to infect dividing and non-dividing cells, efficient, long term expression, non-toxic, low or absent immune response, high packaging capacity, and regulatable expression. At present the most frequently used viruses in rodent research are AAVs and lentiviruses due to their stability, ease of production, low immunogenicity, and low toxicity (Wong et al, 2006; Blessing & Déglon, 2016). The following sections therefore concentrate on these viruses (see Table 1.1 for summary).

1.4.2.1 Adeno-associated viruses

Adeno-associated virus (AAV) was discovered to transfect cells alongside adenoviruses, a family of respiratory viruses, yet is not itself pathogenic (Atchison et al, 1965). AAV belongs to the family parvoviridae, meaning “small”, and the genus dependovirus, as they depend on other viruses for their replication. AAVs are very small with no viral envelope and their genetic material is 4.7kb of linear single-stranded DNA.

AAVs bind to specific glycans on the surface of cell membranes and enter the cell by interaction with a recently characterised receptor (AAVR; Pillay et al, 2016). AAVs have two genes, rep (replication) and cap (capsid), between inverted terminal repeats (ITRs) at each end of their genome. Directed by the ITRs, wild-type AAVs integrate into the genome at a specific site on chromosome 19 by using the host cell polymerase and proteins encoded by the rep gene. During the production of recombinant AAVs (rAAVs) for gene transfer, rep and cap are supplied in trans, i.e. via a separate construct. Helper factors are also required for AAV replication and are supplied by a helper virus such as adenovirus. Lacking the rep gene, rAAVs remain primarily episomal, though may integrate randomly into the genome at a low frequency (McCarty et al, 2004). Circular AAV episomes are formed initially from dsDNA and may then join to form head-to-tail concatemers that are very stable in non-dividing cells and can maintain expression for years (Klein et al, 2002; Carter, 2004).

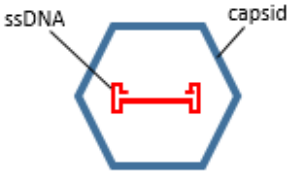
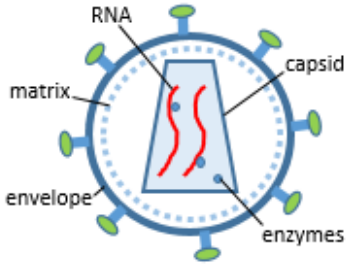
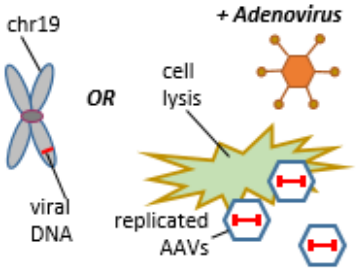
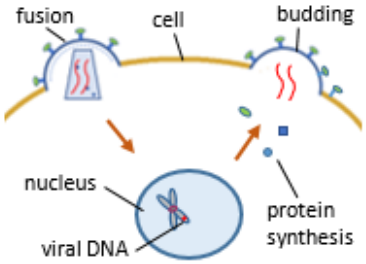
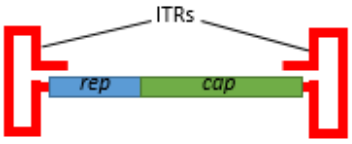
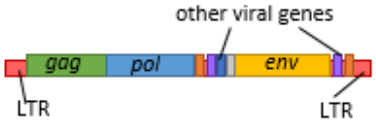
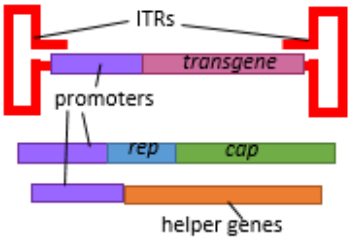
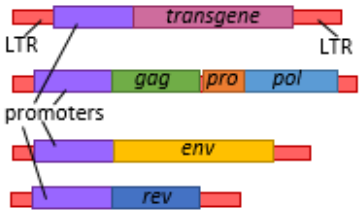
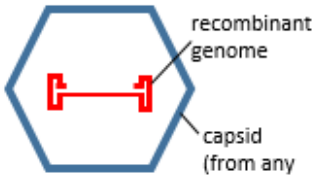
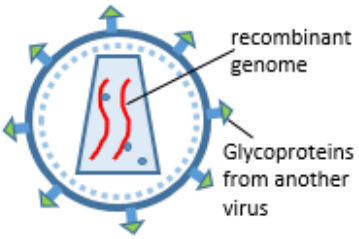
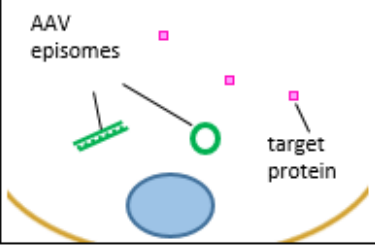
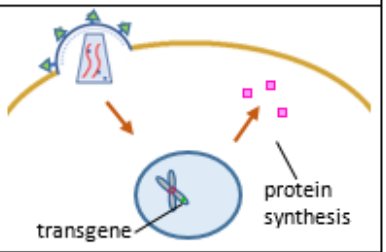
1.4.2.2 Lentiviruses

Lentiviruses are a subtype of retroviruses, named for their slow action, that includes human immunodeficiency virus (HIV). Unlike the simpler oncoretroviruses widely used in early applications, such as the murine leukemia virus (MLV; Hermens & Verhaagen, 1998), lentiviruses have a more complex genome (≈ 9kb) that enables them to cross the nuclear membrane and therefore infect both dividing and non-dividing cells.

Lentiviruses enter the cell by fusion of their envelope with the cell membrane and enter the nucleus by active transport, mediated by interaction of their genes with the nuclear import machinery (Howarth et al, 2010). In the nucleus, lentiviral RNA is reverse transcribed into DNA by the viral reverse transcriptase enzyme and dsDNA is synthesised by the host cell polymerase. Viral DNA is then incorporated into the host chromosomes at varying sites by the viral integrase enzyme. This carries a risk of insertional mutagenesis, which is the disruption of the function or regulation of crucial host cell genes, although incorporation into coding regions is less likely in post-mitotic cells (Bartholomae et al, 2011). Lentiviruses can also form circular episomes and mediate gene expression without integration (Wanisch & Yáñez-Muñoz, 2009). The main genes encoded by wild-type lentiviruses are gag, pol, and env: respectively, the capsid and matrix proteins, reverse transcriptase and integrase, and the viral envelope. Lentiviruses also have two regulatory genes (tat and rev) and additional accessory genes.

Compared to AAVs, the expression of lentiviruses in vivo in non-dividing cells is relatively weak as they usually only express one copy of the viral genome per cell (Nassi et al, 2015) and do not grow to high titres due to toxicity of the viral genes (Howarth et al, 2010).

Table 1.1: Comparison of lentivirus and AAV properties

	AAV	Lentivirus
Virus/virion (outside cell)		
Virus inside cell		
Genome		
Recombinant genome		
Viral vector		
Viral vector in cell		

1.4.2.3 Production of viral vectors safe for in-vivo use

Viral vectors safe for use in-vivo must not cause disease, must prevent reformation or combination with wild type viruses, and should mediate a low immune response. These properties are attained by removing harmful or unnecessary viral genes and supplying genes necessary for viral replication in trans during production. A defective viral vector is one that has had all viral genes removed and retains only the recognition sequences, whereas recombinant vectors retain some of their original genes (Hermens & Verhaagen, 1998).

Transgenes and sequences that aid their expression, such as promoters and enhancers, can be inserted in place of removed viral genes. Exceeding the size of the wild-type virus affects its production and expression, so capacity is limited by the size of the viral genome (Dong et al, 1996; Xu et al, 2001). Lentiviruses are divided between three or more plasmids: the transfer plasmid containing the gene of interest, one or more packaging plasmids, and an envelope plasmid. AAV genes are split between a recombinant AAV plasmid, which contains the transgene and ITRs, and a helper plasmid which codes for the capsid.

The first step in the production of a viral vector, using an existing system, is cloning the target gene, promoter, and other sequences into the viral transfer plasmid and amplifying the construct. Producer cells (e.g. HEK293T cells) are then transfected with the transfer construct and helper plasmids, so that the viral vector can be produced and replicated using the cellular machinery. The viral particles are then extracted and purified and the titre (the concentration of viral particles) is measured.

It is often useful for viral vectors to express a marker gene, under the control of the same promoter as the transgene, to aid the identification of cells that have been transduced by the virus and are expressing the protein of interest. Antibiotic resistance genes are frequently used to select colonies in vitro, while fluorescent proteins are used for visualisation of transduced cells and transgene distribution within individual cells. Fluorescent cells can also be selected using fluorescence-activated cell sorting and this method can be used to titre the virus as well as to study the properties of sorted cells.

1.4.2.4 Modifying viral tropism

The tropism of a virus (the range of host cells that a virus infects) is primarily determined by the proteins on the outer surface of the virus that mediate attachment with different cell types (Bouard et al, 2009). For instance, glycoproteins in the envelope of HIV attach to the glycoprotein CD4, which is present on some immune cells, and so HIV infects only these cells. To manipulate tropism, viruses are often pseudotyped by combining the genome of one virus with the capsid or envelope glycoproteins of another.

AAV tropism differs across AAV serotypes, which have different capsid proteins and ITR sequences. AAVs can be pseudotyped by combining the genome from one serotype with the capsid of another or by creating a hybrid capsid containing the glycoproteins from more than one serotype. Lentiviruses can be pseudotyped with glycoproteins from enveloped viruses, including other retroviruses (Bouard et al, 2009). The most common

pseudotype for lentiviruses is vesicular stomatitis Indiana virus envelope glycoprotein (VSV-G) due to its broad tropism (Blessing & Déglon, 2016) and stability during ultracentrifugation compared to HIV (Naldini et al, 1996). Pseudotyping with VSV-G also reduces the potential for lentivirus to recombine and form wild-type virus (Mundell, 2015).

Within transduced cells, the expression of a transgene depends on its promoter. Promoters can be derived from mammalian genes or from viruses, such as the commonly used promoter cytomegalovirus (CMV; Dull et al, 1998). CMV is a constitutive promoter, meaning that it can direct expression in all cell types. Inducible promoters in contrast are active only under certain conditions and can therefore be used to direct cell-type specificity. More complex patterns of expression are also made possible, including expression in response to light stimuli (optogenetics; Tye & Diesseroth, 2012), tetracycline-inducible systems that enable the activity of certain genes to be switched on or off by the delivery of antibiotics (Drane et al, 2014), and IEG promoters to label neurons activated by an event (Liu et al, 2012b). Site-specific genetic modification in response to these events can be directed by Cre-Lox recombination (Hardy et al, 1997).

Expression of a transgene can also be modified by the insertion of additional regulatory elements such as enhancers, introns to direct splicing (Shevtsova et al, 2005), and a polyadenylation signal to aid mRNA stability and transcription termination (Pfarr et al, 1986; Xu et al, 2001). The Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE; Donello et al, 1998) is frequently used in both lentiviral and AAV constructs as it can increase expression more than tenfold by promoting nuclear export (Xu et al, 2001; Shevtsova et al, 2005; Zufferey et al, 1999).

1.4.3 Summary

Compared to transgenic lines, viral vectors can be produced very quickly, easily, and cheaply and research animals are used more efficiently because not all offspring have the desired expression pattern when breeding transgenic animals. Although using an established transgenic colony is non-invasive, offers greater consistency between subjects, and is less time-consuming compared to individual intracerebral injections, a transgenic line is not readily available for the present application. Viral vectors provided a flexible and efficient method to modify gene expression in the present study as expression of a relatively small construct was required in a spatially-restricted cell population in adult rodents that could be targeted by direct injection.

Lentiviral and AAV vectors have many advantages for the experimental manipulation of neuronal populations when compared with the criteria of Howarth et al (2010), presented in section 1.4.2. Both vectors have a very good safety profile, with low immunogenicity and the removal of most of the viral genome. Both vectors mediate expression in post-mitotic cells such as neurons: lentiviruses primarily by integrating into the genome and AAVs primarily in the form of stable episomes. Lentiviral vectors are preferable for use in dividing cells and cell lines as rAAVs cannot integrate into the genome and so become diluted with multiple rounds of cell division. However, a single

cell can be transduced by multiple AAV particles, whereas cells are usually transduced by only one copy of a lentivirus, and so AAVs can mediate much stronger transgene expression. In some applications this can be a disadvantage: the expression of multiple copies of a transgene may dampen inducible regulation of expression (Wong et al, 2006). AAVs can also be produced to greater titres than lentiviruses and their spread following delivery is greater due to their smaller particles and lack of charge. The time course of expression is similar for both vectors, with slow onset relative to other viral vectors and long-term stability. The expression of transgenes by both types of virus can be modified by promoters and enhancers and is compatible with Cre-Lox recombination, optogenetics, and inducible systems such as tetracycline transactivation. However, the packaging capacity of AAV vectors is very small, limiting the use of larger transgenes and regulatory elements. The packaging capacity of lentiviruses is quite low compared to that of other vectors, but is roughly double that of AAVs.

In summary, the optimal choice of vector depends on the application. If integration into the genome is required or the transgene is quite large, then lentivirus would be the better option, whereas AAV may be preferred when very strong or saturating expression is needed. Both lentiviral and AAV vectors were produced for the present application to optimise expression of the TRAP gene, EGFP-L10a.

1.5 Present work

The aim of the present study was to investigate the genes and regulatory networks involved in associative recognition memory formation by profiling the CA1 neuron translatome during the period following paired viewing of novel and familiar configurations of stimuli (Zhu et al, 1996; Wan et al, 1999). The within-subjects design and passive nature of the task have the advantage of minimising variability in gene expression caused by behavioural and individual differences. Gene expression was compared at four timepoints chosen based on previous studies and research discussed in section 1.2. A matched no-image control condition was used to provide a baseline for gene expression induced by non-visual aspects of the task at the same timepoints.

An additional aim of the project was to develop a viral method of TRAP (Heiman et al, 2008; Doyle et al, 2008) to use in combination with RNAseq to profile the CA1 translatome. Translatome profiling provides a more direct and sensitive measure of protein synthesis compared to profiling the transcriptome and enables investigation of translation-dependent but transcription-independent forms of plasticity.

Previous work has shown that expression of the IEG Fos is upregulated in CA1 by viewing of novel arrangements of stimuli (Wan et al, 1999) and the present study is the first to profile genome-wide expression in CA1 in response to paired viewing of spatial arrangements. We hypothesised that genes would be differentially expressed between the novel, familiar, and no-image control conditions in response to paired viewing, that gene expression would differ between timepoints in all three conditions, that the expression of IEGs would be increased at early timepoints following paired viewing, and that differentially expressed genes would have functions important for synaptic plasticity.

Chapter 2 presents the optimisation of the TRAP method. The main objectives of this chapter were to develop viruses for TRAP that produced strong expression of EGFP-L10a in hippocampal neurons in vivo and in cultures, to optimise the intracerebral injection protocol for the transduction of rat CA1, and to pilot test the purification of ribosome-bound mRNA using both cultured cells and hippocampal tissue transduced by the TRAP viruses.

Chapter 3 presents the development of methods and collection of data for the main experiment combining paired viewing and TRAP. The main objectives of this chapter were to pilot the paired viewing experiment and replicate the finding of increased Fos expression in CA1 in the novel compared to the familiar condition, to extract purified RNA from rodents following paired viewing and prepare libraries for sequencing, to align sequencing data to the rat genome, and to assess the quality of sequencing data and ensure its suitability for downstream analyses.

Chapter 4 presents the analysis of gene expression data from the main experiment combining paired viewing and TRAP, including the optimisation of the analysis pipeline. The main objectives of this chapter were to determine whether there is differential gene expression in the CA1 translatome in response to paired viewing (between the novel and

familiar conditions, between the novel/familiar and control conditions, and between timepoints in each condition) and if so, to identify the differentially expressed genes and their functions. Additional objectives were to analyse the response of IEGs to paired viewing, to identify potential hub genes, and to identify groups of genes with similar expression profiles that may be regulated by shared mechanisms.

Chapter 5 discusses the main findings of the present study and interprets the data in the context of previous research into plasticity processes in CA1.

Chapter 2 Development of a viral method for Translating Ribosome Affinity Purification

2.1 Introduction

This chapter discusses the development of viruses expressing a ribosomal protein tagged with EGFP and the optimisation of viral injections to profile the translatome of CA1 neurons in rats.

2.1.1 Transduction of neurons by viral vectors

The first virus to be used to target the nervous system was herpes simplex virus, a DNA virus that naturally targets neurons and is taken up by sensory nerve terminals (Ho & Mocarski, 1988; Hermens & Verhaagen, 1998). Subsequently, modified forms of adenovirus (Akli et al, 1993), AAV (Kaplitt et al, 1994) and lentivirus (Naldini et al, 1996) to target the nervous system were developed in quick succession (Hermens & Verhaagen, 1998).

Lentivirus and AAV are now commonly used to transduce neurons in vivo with a range of modifications, including pseudotyping to modify tropism and different promoters to modify expression (Blessing & Déglon, 2016). This section provides an overview of factors affecting neuronal expression of these viruses.

2.1.1.1 Targeting neurons by viral tropism

2.1.1.1.1 AAV

The first serotype of AAV to be cloned, AAV2 (Hermonat and Muzyczka, 1984), has been widely used in the CNS but has a relatively low efficiency compared to pseudotypes of AAV2 with the capsid proteins of other serotypes (Burger et al, 2004; Cearley et al, 2008; McFarland et al, 2009; Van der Perren et al, 2011). Indeed, most AAV serotypes transduce neurons preferentially or exclusively over other CNS cell types, with varying specificity and efficiency (Taymans et al, 2007; Cearley et al, 2008).

Capsid serotypes 1, 2, 5, 6, 8 and 9 have all been found to transduce neurons, but their affinities for neurons differ between regions (Aschauer et al, 2013). In vivo in the rat hippocampus, a succession of studies has found that serotypes 1 and 5 preferentially transduce CA1 and CA3 PNs compared to AAV2, which has affinity for the dentate hilus (Burger et al, 2004), and the closely related serotypes AAV8 (Klein et al, 2006) and AAV9 (Klein et al, 2008) may further increase the efficiency of neuronal transduction (see Cearley et al (2008) for relationships between serotypes). AAV9 was found to transduce CA1 neurons preferentially over the rest of the hippocampus compared to AAV8 and other serotypes, quantified by both Western blot and biophotonic imaging (Klein et al, 2008). In concordance with other studies using AAV9, there was little to no transduction of glia (Cearley & Wolfe, 2006; Cearley et al, 2008). However, glial expression was observed in a subsequent experiment utilising the same AAV9 vector (Dayton et al,

2012b). In mice, AAV9 had the greatest spread compared to serotypes 7, 8 and rh10 and expression in CA1 neurons reached saturating levels (Cearley & Wolfe, 2006). Due to its high efficiency, AAV9 has also been used for global transduction of the CNS by intravenous (Dayton et al, 2012a) and intracerebroventricular (McLean et al, 2014) delivery to neonatal mice. Novel serotypes may further improve neuronal transduction (Cearley et al, 2008), though with large volumes of AAV (1-4µl) expression differences between serotypes may be masked by the reaching of a saturation point (Scheyltjens et al, 2015).

AAV can be both anterogradely and retrogradely transported with varying efficiency depending on serotype (Castle et al, 2014; Scheyltjens et al, 2015). Differences in the frequency of transport may underlie variation in spread between AAV serotypes, though it has been suggested that these differences are due to variation in vector uptake rather than in transport mechanisms (Castle et al, 2014). Cearley & Wolfe (2006) observed particularly strong transport of AAV9 to other regions, including the contralateral hippocampus, as well as greater spread compared to other serotypes tested. Another factor that may contribute in practice to serotype differences in spread is that some serotypes naturally grow to higher titres than others (Holehonnur et al, 2014), though titre is usually normalised between serotypes in comparison experiments.

In rat primary hippocampal cultures using a CMV promoter, AAV2 and AAV9 have been found to have relatively weak expression in neurons over astrocytes compared to serotypes 1, 7 and 8 (Royo et al, 2008; Howard et al, 2008). As these results suggest, transduction patterns for AAV are often substantially different in cultured neurons compared to in vivo (Shevtsova et al, 2005; Klein et al, 2006). This has been attributed to the time required for AAVs to form stable episomes (Carter, 2004). In dividing cells, rAAVs become diluted over time as they are not integrated into the genome.

In vivo, AAV episomes are highly persistent and expression can be detected more than two years post-injection (Klein et al, 2002). The onset of expression takes approximately 5-7 days (Kugler et al, 2003), though this varies between serotypes (Reimsnider et al, 2007).

2.1.1.1.2 Lentivirus

To transduce neurons, lentiviruses are commonly pseudotyped with VSV-G (Naldini et al, 1996; Blessing & Déglon, 2016). Although the tropism of VSV-G is broad, it has been reported to show preferential transduction of neurons in vivo (Naldini et al, 1996; Mazarakis et al, 2001). A wide range of pseudotypes for lentivirus exists, some of which have been tested in the CNS (Wong et al, 2006; Bouard et al, 2009).

A comparison of transduction efficiency in the striatum between different pseudotypes of the lentivirus equine infectious anemia virus found that VSV-G (Indiana strain) and rabies virus both transduced neurons with approximately 90% specificity (Wong et al, 2004). Another strain of VSV-G, Chandipura, was substantially less selective for neurons. Mokola virus, a member of the same genus as rabies virus, has been suggested as an alternative pseudotype for the transduction of neurons due to its neural tropism (Desmaris et al, 2001; Watson et al, 2002), but other studies have found that

transduction with this pseudotype is relatively weak (Desmaris et al, 2001; Wong et al, 2004). An alternative approach is to engineer pseudotypes to target specific molecules on the surface of particular cell types (Bouard et al, 2009).

Rabies virus glycoprotein is exclusively and efficiently retrogradely transported (Astic et al, 1993) and is therefore particularly useful where there is a need to select neurons based on their connectivity (Nassi et al, 2015). VSV-G is anterogradely transported (Beier et al, 2011), but a hybrid pseudotype combining rabies and VSV-G (fusion glycoprotein C-type) is more efficiently retrogradely transported than rabies virus and has reduced affinity for glia (Kato et al, 2011).

The onset of expression for lentivirus is slow and although expression is observed within one week, it may take much longer to reach a plateau (Hioki et al, 2007). Expression in the rodent brain can be sustained for at least one year (Wong et al, 2006).

2.1.1.2 Promoters and enhancers

Different promoters and enhancers can be cloned into viral vectors to control the cell-type specificity and efficiency of transgene expression. A constitutive promoter such as CMV or chicken beta-actin (CBA) is often used for viral vectors and so cell-type specificity is controlled primarily by viral tropism in these cases. Alternatively, promoters specific to neurons or excitatory neurons can be used, within the limited capacity of AAV and lentivirus.

2.1.1.2.1 AAV

The CMV sequence used in viral vectors includes an enhancer and a 5' UTR sequence in addition to the promoter itself. CMV initially produces strong expression using AAV but in some tissues this diminishes over time as the promoter is silenced. In the hippocampus, AAV expression with a CMV promoter has been shown to peak approximately 3 weeks post-injection and decline within 2-3 months (Klein et al, 1998; Gray et al, 2011). Alternatives are therefore needed for longer term applications.

The endogenous constitutive promoter CBA is not switched off over time (Gray et al, 2011) and a combination of the CBA promoter with a CMV enhancer can mediate strong and stable neuronal expression (Klein et al, 2002). The CBA and hybrid promoters are more than twice the size of the CMV promoter, but shortened forms are also effective (Gray et al, 2011).

A range of cell type specific promoters have also been used to target expression with AAVs. Global delivery of AAV9 to the CNS in neonates produces non-specific expression in the CNS (Dayton et al, 2012a), but with a synapsin promoter, this expression can be targeted to neurons (McLean et al, 2014). The neurone specific enolase (NSE) promoter also drives expression in neurons (Xu et al 01), though is particularly large at 1800bp, and a CaMKii promoter can be used to target excitatory neurons (Dittgen et al, 2004; Scheyltjens et al, 2015). However, promoter activity is subject to viral tropism. For instance, expression in the hippocampus using AAV2 remained neuron-specific even using the glial promoter GFAP (Xu et al, 2001).

To maximise transgene capacity, a fragment of a promoter can be sufficient for directed expression. A 480bp fragment of synapsin and a ~300bp fragment of NSE each drive neuron specific expression with a strength comparable to or greater than CMV (Shevtsova et al, 2005; Xu et al, 2001). An even smaller 229 bp methyl-CpG-binding protein-2 (MeCP2) promoter drives relatively specific but weak neuronal expression (Gray et al 2011). There is usually a trade-off between the specificity and size of a promoter, though this is not always the case: interestingly, one study in mouse V1 found that with a 1.3kb CaMKii promoter the proportion of transduced neurons that were excitatory was very similar to CMV (approximately 80%), yet with a shorter 0.4kb CaMKii promoter 95% of transduced neurons were excitatory (Scheyltjens et al, 2015).

2.1.1.2.2 Lentivirus

The expression of wild type lentivirus is controlled by a promoter in the long terminal repeat (Zufferey et al, 1998). To increase safety for in vivo use, the vector is commonly made self-inactivating by a deletion in this promoter (Zufferey et al,1998; Nassi et al, 2015). This prevents formation of replication-competent virus during production and recombination with wild type lentiviruses in vivo. Instead, lentiviral vectors commonly use constitutive promoters that drive strong, ubiquitous expression such as CMV and phosphoglycerokinase (Dull et al, 1998; Wong et al, 2006).

In vivo, lentivirus pseudotyped with VSV-G has mainly neural tropism with constitutive promoters (Naldini et al, 1996; Jakobsson et al, 2003) and in rodent hippocampus, the CMV promoter has been shown to drive selective expression in CA1 pyramidal neurons over other cell types (Kuroda et al, 2008; van Hooijdonk et al, 2009). In contrast to AAV, lentivirus with a CMV promoter can sustain stable transgene expression in neurons for at least six months (Blömer et al, 1997). The constitutive promoters CAG (a CMV/beta-actin hybrid) and elongation factor 1a (EF1a) also drive strong selective expression in neurons and the results of Jakobsson et al (2003) in rat striatum suggest that CAG is a stronger promoter than CMV while EF1a is more selective for neurons.

The neuron specific promoters NSE and synapsin can be used to drive expression in neurons with very high specificity (NSE: Jakobsson et al, 2003; Tian et al, 2009; Synapsin: Dittgen et al, 2004; Hioki et al, 2007). However, expression is generally found to be much stronger with CMV than with neuron and excitatory neuron specific promoters (Hioki et al, 2007; Kuroda et al, 2008). The addition of a CMV enhancer to cell type specific promoters can increase transgene expression by 2-4 times, though with some loss of specificity (Hioki et al, 2007).

2.1.1.3 Summary

Lentiviral and AAV vectors can both be used to transduce neurons in the rodent hippocampus, with strength and selectivity dependent on many factors.

The proteins on the virus outer surface mediate cell entry and therefore are key to tropism, while expression within a transduced cell is orchestrated by promoters and enhancers. Increasing the selectivity of transduction is preferable to increasing the selectivity of expression as this reduces the risk of off-target effects (Anliker et al, 2010)

as well as being more efficient. The choice of capsid or envelope pseudotype therefore supersedes the choice of promoter in vector design (Shevtsova et al, 2005; Bouard et al, 2009).

Of the several AAV serotypes with selectivity for neurons, AAV9 has been widely reported to have particularly strong expression in the CNS in vivo (Klein et al, 2008; McLean et al, 2014) and is readily transported between neurons in both directions (Castle et al, 2014) which may contribute to its increased spread. There are inconsistencies in the literature regarding whether AAV9 has neuronal specific tropism, but specific expression can be directed using short fragments of promoters such as synapsin and CaMKiiα. AAV shows substantial differences in expression between neurons in culture and in vivo.

Lentivirus has preferential tropism for neurons with VSV-G and rabies virus pseudotypes. Rabies is strongly transported in the retrograde direction and so is ideal for projection mapping. Lentivirus has a greater capacity for specific promoters and enhancers than AAV, but its expression with inducible promoters can be weak. The choice of promoter may be driven by whether specificity or strength of expression is more important, though strong enhancers may compensate for a weaker promoter.

Regardless of the vector chosen, it is important to investigate its expression before use as results in the literature are inconsistent and tropism may be drastically altered by subtle differences between preparations: one study found that using a different method of purification changed the tropism of AAV8 (but not other serotypes) from neurons to astrocytes (Klein et al, 2008).

2.1.2 Present work

This chapter discusses the development of methods for the viral delivery of EGFP-tagged L10a to profile mRNAs undergoing translation in CA1 neurons.

The key characteristics of the virus needed for the present application are ability to transduce post-mitotic cells, induction of strong expression, spread throughout the CA1 pyramidal layer but ideally not beyond, and expression that remains stable for at least two months for recovery and the conduct of behavioural experiments. Ribosomal proteins in rat brain have a half-life of 12 days (von Hungen et al, 1968) and so an incubation period is required for the transgene to integrate into functional ribosomes. Strong expression of EGFP-L10a is essential to obtain sufficient ribosome-bound mRNA for TRAP. Injections were preferentially targeted to dorsal hippocampus due to its greater involvement in spatial memory (Fanselow & Dong, 2010). Spread of the virus outside the CA1 subregion can be compensated for by a subregion-specific dissection using a method that has previously been used for the transcriptional profiling of hippocampal subregions (Zhao et al, 2001).

Lentiviruses and AAVs expressing EGFP-L10a under the control of CMV and CaMKiiα promoters were developed and viral injections were optimised for the transduction of CA1 neurons in adult rats by comparing the viruses and different co-ordinates. The TRAP protocol was then tested on CA1 tissue dissected from subjects that had received

optimised injections and the number of pooled hippocampi necessary to obtain sufficient RNA for sequencing was determined. Lentivirus expressing EGFP-L10a was also used to transduce rat hippocampal neurons in vitro and to produce a stable cell line suitable for TRAP.

2.2 Methods

2.2.1 Virus production

2.2.1.1 Cloning

A plasmid containing the EGFP-L10a gene was provided by collaborators at Takeda Pharmaceutical Company Ltd (Cambridge, UK). The EGFP-L10a sequence was cloned into a lentiviral backbone vector plasmid (pRRL) and an AAV backbone vector (supplied by Younbok Lee, King's College London), both initially expressing EGFP under a CMV promoter. Additionally, a CaMKiiα promoter from the plasmid pLenti-CaMKiiα-eNpHR3 was cloned into the pRRL vector expressing EGFP to compare expression patterns. Plasmid maps were visualised using Serial Cloner (Serial Basics, France).

2.2.1.1.1 Method 1

The pRRL vector plasmid was digested sequentially with the restriction enzymes Age1 and Sal1 to cut the plasmid at both ends of the EGFP sequence and dephosphorylated using alkaline phosphatase. The vector was separated from the EGFP sequence by gel electrophoresis and cut from the gel under UV light in a GelDoc-It Imaging System (Ultra-Violet Products Ltd, Cambridge, UK). The DNA was purified using QIAquick gel extraction kit (QIAGEN, Germany). The insert was obtained by amplifying the EGFP-L10a gene by polymerase chain reaction (PCR) using FastStart High Fidelity PCR System (Roche) with two primers (synthesised by Sigma Aldrich, MO, USA), each matching one end of the insert sequence and one end of the digested vector. A nanophotometer (NanoPhotometer Pearl, Implen, Germany) was used to measure the DNA concentration of the vector and insert to calculate the volumes of each solution to use in the ligation reaction. The vector and insert were incubated with DNA ligase for 15 mins at room temperature using Roche rapid DNA ligation kit (Roche, Switzerland). The plasmids were transformed into DH5α cells (Invitrogen, MA, USA) by heatshocking for 20s at 42°C and then incubated without antibiotic for 1h at 37°C. The cells were spread on LB agar plates containing ampicillin and incubated overnight at 37°C.

2.2.1.1.2 Method 2: Blunt ended ligation

The pRRL vector and pLenti-CaMKiiα plasmid were digested using the restriction enzymes Cla1 and Pac1 respectively to cut the plasmids at the beginning of the promoter sequence. The cut ends were then blunted using Mung bean nuclease, with PCR clean-up (QIAquick PCR purification kit, QIAGEN, Germany) performed between steps. Both polynucleotides were then digested with BamH1 to cut the plasmids at the end of the promoter and the vector was dephosphorylated using alkaline phosphatase. The vector and insert were both separated by electrophoresis and purified. The remainder of the procedure was as described for method 1.

2.2.1.1.3 Method 3: InFusion cloning

This method used the InFusion cloning kit (Clontech Laboratories, CA, USA). The pRRL and AAV vector plasmids were linearised by digestion with the restriction enzymes Age1 and Sal1-HF, and Age1 and BsrG1 respectively to cut the plasmid at both ends of the EGFP sequence. Primers to match the ends of each vector and the EGFP-L10a insert were designed using the InFusion online PCR primer design tool and then synthesised

(Sigma-Aldrich). The inserts were amplified by PCR using the custom primers. The desired vector and insert sequences were extracted by electrophoresis and purified. The insert and vector were then incubated for 15min at 50°C with 5x InFusion HD Enzyme Premix (Clontech). The DNA was transformed into Stellar Competent cells (Clontech) by heatshocking for 45s at 42°C and then incubated without antibiotic for 1h at 37°C. The cells were spread on LB agar plates containing ampicillin and incubated overnight at 37°C.

2.2.1.2 Amplification

Individual colonies were picked from the agar plates and grown in 5ml LB and ampicillin in a shaking incubator for 12-16 hours at 37°C. The cells were then mini-prepped to extract the DNA. To test for the presence of the target plasmid, each sample was incubated with a restriction enzyme with one target site in the insert and another in the plasmid for 2-3 hours and then separated by electrophoresis. Two samples containing the target plasmid were transformed into Stbl3 competent cells (Invitrogen) and incubated on agar plates at 37°C overnight. Colonies from the plates were grown overnight in LB and ampicillin in a shaking incubator at 37°C and the mini prep, restriction digest and electrophoresis were repeated the following day to confirm presence of the target plasmid.

1ml of cells was added to a sterile 2L conical flask containing 500ml LB media and ampicillin and grown overnight at 37°C in a shaking incubator. The cells were then maxi-prepped to extract the DNA. The concentration was measured using a nanophotometer.

To confirm the identity of the plasmid produced, the restriction digest and electrophoresis were repeated and samples were sequenced by Source Bioscience (Cambridge, UK) using a combination of the custom-made PCR/InFusion primers and the standard primers CMV forward, EGFP C-terminal forward and EGFP N-terminal reverse. The sequencing data was analysed by alignment with the target construct using the online tool Clustal Omega Multiple Sequence Alignment (European Bioinformatics Institute, UK; Sievers et al, 2011). For sections of DNA sequenced with reverse primers, sequences were reverse complemented prior to alignment. Errors in the sequencing data compared to the target constructs were counted after discarding unaligned portions at the beginning and end of each sequence. Total alignment was calculated by subtracting errors and unaligned sections from the overall sequence length.

2.2.1.3 Lentivirus production

2.2.1.3.1 Cell culture

HEK293T cells were maintained in media (Dulbecco's Modified Eagle Medium with 10% foetal bovine serum, 2mM L-Glutamine, 50µg/ml penicillin/streptomycin and 5ml non-essential amino acids) in an incubator with 5% CO2 at 37°C. Cells were passaged every 3-4 days.

2.2.1.3.2 Virus production

Lentivirus was produced using the third-generation system developed by Dull et al (1998). HEK293T cells were plated and co-transfected the following day with the target

plasmid DNA and the packaging plasmids pMPDLg-Prpe, pRSv-Rev and pMD2-VSVG. Media was replaced after 12 hours with media containing sodium butyrate to upregulate transcription. Media containing virus was harvested seven hours after the media change. A second harvest was taken 24 hours after the first. The harvested media was spun to remove cell debris and the resultant supernatant was filtered then spun overnight at 6000g at 4°C in a Sorvall 300 centrifuge (Thermo Scientific, MA, USA). The viral pellet was resuspended in cold PBS and spun at 20,000rpm for 90 minutes at 4°C in a Beckman Ultracentrifuge with an SW40 rotor (Beckman Coulter, CA, USA). The viral pellet was resuspended in tromethamine, sodium chloride, sucrose and mannitol (TSSM) on ice over several hours and stored in aliquots at -80°C.

2.2.1.3.3 Titring

The titre was measured by transducing HEK293T cells with both viruses and a virus of known titre at 1 in 1000 and 1 in 5000 dilutions. The cells were grown for one week in 12 well plates alongside untransduced cells and passaged every 2-3 days. Two biological replicates were made for each condition. The cells were then transferred to six well plates and grown for three days, before samples were harvested and stored at -20°C. The genomic DNA was extracted using Wizard Genomic DNA purification kit (Promega, WI, USA). The concentration of DNA was then measured using a nanophotometer and samples were diluted to 19.5-21.0ng/µl. Quantitative PCR (qPCR) was conducted in triplicate on DNA from each sample alongside serial dilutions of the pRRL backbone from 10^{-4} to 10^{-8} to produce a standard curve.

2.2.1.3.4 AAV production

Two AAV9 viruses expressing CMV-EGFP and CMV-EGFP-L10a were produced and titred by qPCR at King's College London by Younbok Lee and Doyoung Lee. The titres were $2.4x10^{12}$ particles/ml and $7.76x10^{11}$ particles/ml respectively.

2.2.2 Primary culture and transduction by lentivirus

2.2.2.1 Neuronal culture

Primary hippocampal cultures were obtained from embryonic day 18 (E18) Wistar rats. Cells were plated on poly-D-lysine coated coverslips in 24 well plates at a density of 65,000 cells/well. Cells were maintained in media (neurobasal media containing 2% B27, 100µg/ml penicillin/streptomycin and 0.5mM L-glutamine) in an incubator with 5% CO2 at 37°C and fed every 3-4 days by 50% media exchange. Transduction with lentivirus was performed at three days in vitro (div).

2.2.2.2 Fixation and visualisation

Cells were fixed with 4% PFA at 15 div. Coverslips were fixed to slides with mounting medium and visualised with a Leica fluorescent microscope.

2.2.2.3 Immunocytochemistry

PFA-fixed cells were blocked in 10% donkey serum and 0.1% Triton X detergent in PBS for 1 hour. Wells were then incubated with primary antibody in 1% donkey serum and PBS at 4°C overnight. Primary antibodies used were: anti-βIII-tubulin (1:500, Millipore AB9354), anti-GFAP (1:500, Dako Z0334), anti-GAD67 (1:100, Abcam ab26116; GAD:

glutamate decarboxylase), anti-CaMKiiα (1:500, Millipore 05532) and anti-GFP (1:2000, Roche 1181446000). Wells transduced with CMV-EGFP-L10a and incubated with βIII-tubulin and GFAP were simultaneously double labelled with anti-GFP to enhance the EGFP signal in cells expressing EGFP-L10a. Wells were then washed 3 times with PBS and incubated with a 1:500 dilution of fluorescent-conjugated secondary antibodies (CY2 for anti-GFP, otherwise CY3) raised in donkey for 1-2hrs. Wells were washed again and incubated with a 1:1000 dilution of Hoechst for 15 minutes to stain nuclei. Coverslips were fixed to slides with mounting medium and visualised under a Leica fluorescent microscope at 10, 20, 40, and 63x magnification. Images were captured in greyscale and colour was applied during processing in Leica Application suite.

2.2.3 Viral injections

2.2.3.1 Subjects

Adult male Lister Hooded rats (Harlan, UK) aged 3-4 months and weighing >300g were used for viral injections. Animals were maintained on a 12-hour reverse light cycle with experiments performed during the dark (active) phase. All procedures were performed under licence in accordance with the UK Animals (Scientific Procedures) Act 1986.

2.2.3.2 Surgery

Subjects were anaesthetised in a plexiglass box with 4% isoflurane in gas containing 95% O2 and 5% CO2 at a flow rate of 1L/min. Once a sufficient depth of anaesthesia was reached (loss of foot and tail pinch reflexes and steady breathing rate), animals were shaved and placed in a stereotactic frame with a face mask on top of a heated pad. The incisor bar was set at -3.3mm for a flat skull. Iodine or chlorhexidine and lidocaine were applied to the scalp to sterilise and anaesthetise the area respectively.

An incision was made using a scalpel to expose the skull surface. Gas flow rate was then reduced to 0.6L/min and isoflurane to 3.5%, gradually lowering to 2% over the course of surgery. The co-ordinates of bregma and lambda were measured to confirm a flat skull (heights within ±0.5mm). Bilateral injections were made at a range of co-ordinates to optimise targeting of CA1. The first set of surgeries used an anterior-posterior (AP) co-ordinate of -5.3mm, latitude (lat.) co-ordinates between ±3.0 and ±4.0mm, and depth co-ordinates between -3.0 and -4.0mm (see Table 2.5). Optimised co-ordinates from these surgeries (AP -5.3, lat. ±3.5, depth - 3.0mm) were subsequently compared to another set of co-ordinates chosen to target dorsal CA1 (AP -3.3, lat. ±2.0, depth -2.0mm; Wong et al, 2005). The latter co-ordinates were used for further surgeries. The positions of the injection sites were calculated from bregma, marked, and drilled freehand. A needle was used to puncture the dura. The co-ordinates were measured again with a 10µl Hamilton syringe and virus was injected at a rate of 0.2µl/min, then left in place for a further 10 minutes. Following each injection, 2.5ml of saline was delivered subcutaneously. To improve surgical technique following the first set of injections, the needle was rinsed with ethanol and autoclaved H_2O between injections and the hemisphere receiving the first injection was counterbalanced between subjects in all subsequent surgeries. The wound was then sutured and covered in antiseptic powder. An intramuscular injection of 0.05ml analgesic (Vetergesic) was administered

for pain relief and eye drops were applied. Animals were then removed to a warm recovery chamber until consciousness and righting reflex were regained. The recovery of subjects was monitored for one week following surgery or until their weights had stabilised.

2.2.3.3 Histology

Subjects were transcardially perfused three weeks post-surgery with 0.1M phosphate buffer (PB), followed by 4% paraformaldehyde (PFA). Brains were post-fixed in PFA for 24 hours and then transferred to 30% sucrose solution for 48 hours. Brains were snap frozen in liquid nitrogen and 40µm coronal sections were taken using a cryostat in a darkened room to prevent bleaching of EGFP. Sections were mounted immediately onto gelatin-coated slides.

Glass coverslips were applied using DAPI (4',6-diamidino-2-phenylindole) fluorescent mounting medium and slides were visualised using a Leica fluorescent microscope (Leica, Germany). Colour images were captured and processed in Leica Application Suite. Overlays of images from the green and blue channels also included an image from the red channel with the same settings as for the green channel (EGFP) to cancel out background fluorescence.

2.2.4 Translating Ribosome Affinity Purification (TRAP)

2.2.4.1 Cultured cells

HEK293T cells, cultured as described in 2.2.1.3.1, were transduced with lentivirus CMV-EGFP-L10a to produce a stable cell line expressing the TRAP gene. These cells were grown on a 15cm diameter plate.

Protocol was carried out as in Heiman et al (2014; see Appendix for list of solutions used for TRAP). Cells were incubated with cycloheximide to stall translation of mRNA, then dissociated with a cell lysis buffer and homogenised. Lysates were centrifuged and 1,2-diheptanoyl-sn-glycero-3-phosphocholine (DHPC) was added to maintain the integrity of the ribosomes. Lysates were then spun at high speed (16,500rpm) for 25 minutes. The supernatant was incubated with protein-L conjugated magnetic beads coated in a mixture of two primary antibodies to EGFP derived from goat (Memorial Sloan-Kettering Monoclonal Antibody Facility; clone names: Htz-GFP-19F7 and Htz-GFP-19C8; Doyle et al, 2008, Heiman et al, 2014) at 4°C for at least 2 hours to precipitate EGFP-tagged ribosomes and associated mRNA. The beads were washed in a 0.35M KCl buffer and RNA was purified using the Absolutely RNA Nanoprep kit (Agilent, CA, USA), alongside samples taken before and after immunoprecipitation (the input and unbound samples, respectively). RNA concentration was measured using a nanophotometer.

2.2.4.2 Rat brain tissue

Rats were sacrificed by concussion and death was confirmed by cervical dislocation. The brain and hippocampi were rapidly dissected on ice in dissection buffer containing cycloheximide to stall translation of mRNA. CA1 was then dissected using a small scalpel to separate the subregion from the rest of the hippocampus as described in Lein et al (2004) in dissection buffer containing 10% RNasin in addition to cycloheximide. The

remainder of the hippocampus was transferred to RNAlater for storage at -80°C. The first set of pilot dissections were performed using an alternative CA1 dissection method described by Sultan (2013). All tools, filter paper and homogenisers were cleaned with RNaseZap, rinsed with RNase free water and autoclaved prior to use.

CA1 samples were pooled and homogenised in lysis buffer containing RNasin, SuperASEin and cycloheximide using manual glass homogenisers. TRAP was then carried out as described above for cultured cells from the lysate centrifugation step. RNA concentration was measured using a nanophotometer and RNA integrity was measured using the Agilent 2200 TapeStation (Agilent, CA).

2.3 Results

2.3.1 Preparation of AAV and Lentiviruses for TRAP

2.3.1.1 Cloning of EGFP-L10a and promoters into viral backbone plasmids

The TRAP gene, EGFP-L10a, was cloned into lentiviral and AAV backbone plasmids with CMV and CaMKiiα promoters. Table 2.1 shows the list of plasmids prepared.

Table 2.1: Viruses produced and viral backbone plasmids cloned in house

Virus type	Promoter	Gene	Plasmid	In house?	Virus	In house?
Lentivirus	CMV	EGFP	Y	N	Y	N
Lentivirus	CMV	EGFP-L10a	Y	Y	Y	Y
Lentivirus	CaMKii	EGFP	Y	Y	Y	Y
AAV	CMV	EGFP	Y	N	Y	N
AAV	CMV	EGFP-L10a	Y	Y	Y	N
AAV	CaMKii	EGFP	Y	Y	N	-
AAV	CaMKii	EGFP-L10a	Y	Y	N	-

For each new construct, a restriction enzyme with one target site in the insert and another in the vector was found by examining maps of each plasmid on Serial Cloner (Serial Basics, France) and the lengths of the expected products were calculated, as listed in Table 2.2.

Table 2.2: Restriction digests to confirm cloning of viral backbone plasmids

Backbone plasmid	Promoter	Gene	Restriction enzyme	Expected lengths
Lentivirus	CMV	EGFP-L10a	EcoR1	6849, 1513
Lentivirus	CaMKii	EGFP	Stu1	5254, 2872
AAV	CMV	EGFP-L10a	Xho1	3911, 1754
AAV	CaMKii	EGFP	PvuII	2364, 1766, 1166
AAV	CaMKii	EGFP-L10a	Xho1	3906, 2328

Restriction digests were carried out on multiple samples after transformation of the plasmid into DH5α cells to determine the success of the ligation reaction. The digests were repeated after transformation into Stbl3 cells and after maxiprep. The resulting

solutions were separated by electrophoresis alongside a 1KB DNA ladder and imaged under UV light. The results were then compared to the expected lengths. Figure 2.1A shows gels from the digests of samples for each target construct following the ligation reaction and transformation and Figure 2.1B shows gels from digests following maxiprep of DNA. Asterisks indicate the mini-prep samples that were selected for maxiprep.

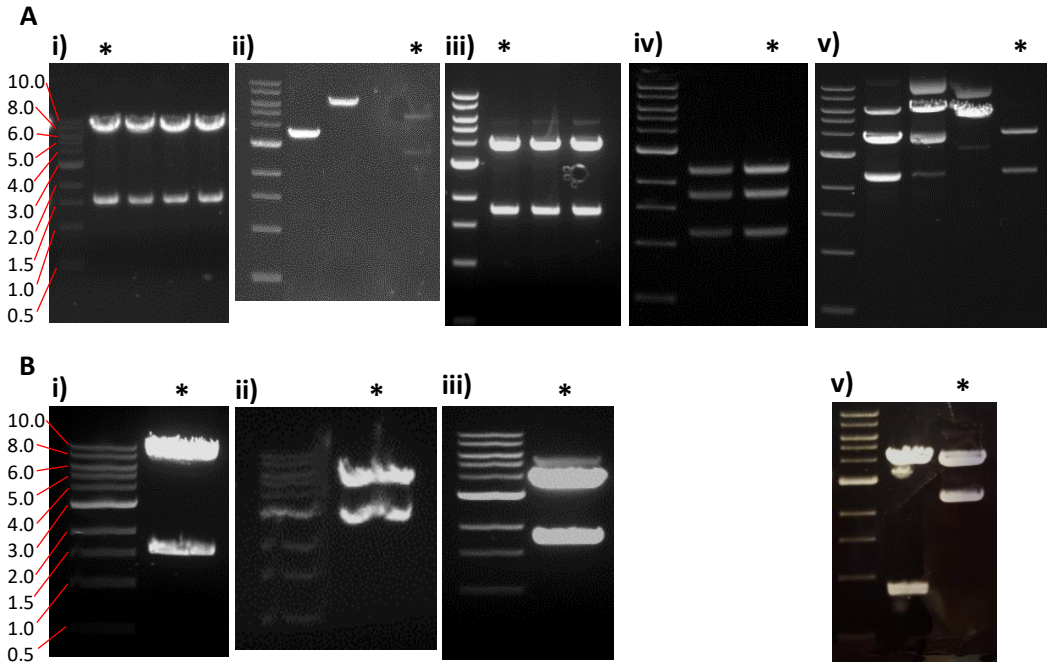


Figure 2.1: Bands of DNA from restriction digests of viral plasmids separated by gel electrophoresis and visualised under UV light with a 1KB DNA ladder for reference. Target samples are marked by asterisks. A: After miniprep. B: After maxi-prep. i) pRRL CMV-EGFP-L10a; ii) pRRL CaMKiiα-EGFP; iii) AAV backbone CMV-EGFP-L10a; iv) AAV backbone CaMKii-EGFP; v) AAV backbone CaMKii-EGFP-L10a.

Samples of the plasmids were sent for Sanger sequencing (Source Bioscience, Cambridge, UK) using a range of primers. Table 2.3 shows the alignment statistics for each sequencing reaction. Alignment scores of 90-95% were obtained for most target and primer combinations. A lower score (75%) was obtained sequencing pRRL CMV-EGFP-L10a with the InFusion forward primer used to clone the plasmid, but this was due to a large number of unsequenced bases rather than errors. Sequencing of AAV CMV-EGFP-L10a with the CMV forward primer was not successful.

Table 2.3: Sequencing results for cloned plasmids prior to virus production.The table shows primers used for sequencing, length of sequence returned, lengths of unaligned sections, unsequenced bases (Ns), and insertion (ins.), deletion (del.) and substitution (sub.) errors compared to the target sequences. Total alignment was calculated by subtracting errors and unaligned sections from the overall sequence length.

Target plasmid	Primer	Sequence length	Unaligned (start)	Errors				Unaligned (end)	N	Total % alignment
				Ins.	*Del.*	*Sub.*	***Total***			
pRRL CMV-EGFP-L10a	EGFP C Fwd	1348	7	1	41	22	64	5	0	**94.36**
	CMV Fwd	1196	17	19	0	18	37	0	0	**95.48**
	InFusion Fwd	1236	29	0	0	0	0	274	275	**75.49**
	InFusion Rev	1002	34	3	1	1	5	43	149	**91.82**
AAV CMV-EGFP-L10a	EGFP C Fwd	1266	7	1	2	21	24	86	0	**90.76**
	EGFP N Rev	935	15	3	1	5	9	36	0	**93.58**
	CMV Fwd	156	156	-	-	-	-	-	120	**0.00**
pRRL CaMKiiα-EGFP	EGFP N Rev	1406	50	0	8	50	58	12	0	**91.47**

2.3.1.2 Virus production and titre quantitation by qPCR

The titres of the lentiviruses CMV-EGFP-L10a and CaMKiiα-EGFP were determined by qPCR of genomic DNA extracted from HEK293T cells transduced by virus.

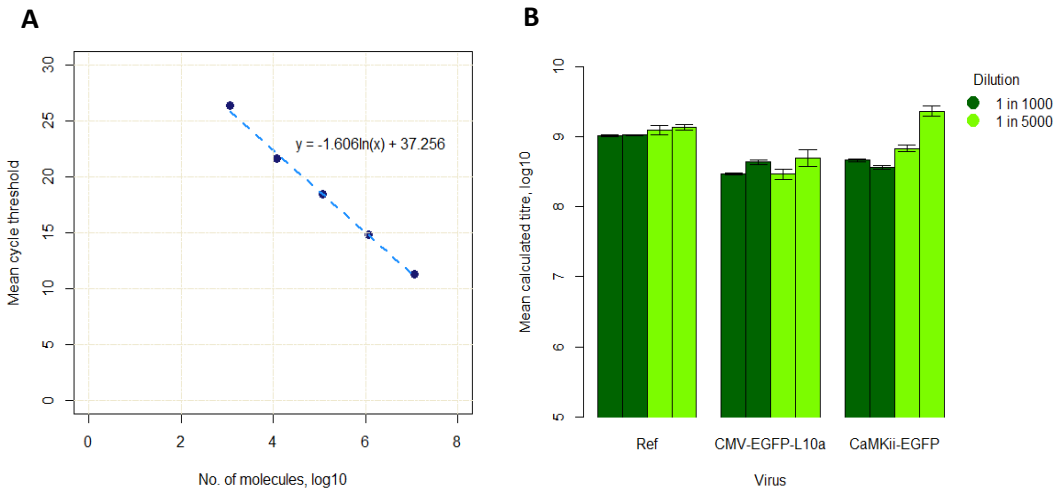


Figure 2.2: qPCR data for titre quantitation of lentiviruses CMV-EGFP-L10a and CaMKiiα-EGFP. A: Standard curve used to calculate titre of unknown samples. B: Titres of N=2 biological replicates of two dilutions (1 in 1000 and 1 in 5000) of CMV-EGFP-L10a, CaMKiiα-EGFP and a reference virus. The bar chart shows the mean and standard error of the mean (SEM) of calculated titres from three technical replicates and is plotted on a log_{10} scale.

The titre of the new viral preparations was calculated using the standard curve shown in Figure 2.2A, which was produced from dilutions of pRRL backbone plasmid, and a virus of known titre. Figure 2.2B shows the titres of two biological replicates of two dilutions of each virus from qPCR plates prepared in triplicate. One sample, a 1 in 5000 dilution of CaMKiiα-EGFP, was excluded from the subsequent calculation of average titre as the concentration of DNA was too low to detect by nanophotometer, though the titre estimated using this sample was higher than the other estimates by a factor of ten. Except for this sample, the estimated titres were consistent between technical and biological replicates and dilutions. The average titres were 3.79×10^8 particles/ml for CMV-EGFP-L10a and 5.50×10^8 particles/ml for CaMKiiα-EGFP.

2.3.2 Expression of lentiviruses in primary hippocampal culture

2.3.2.1 Imaging of hippocampal primary cultures

Primary hippocampal cultures from E18 Wistar rats were obtained, transduced with the lentiviruses CMV-EGFP-L10a and CaMKiiα-EGFP at 3 div, and fixed at 15 div.

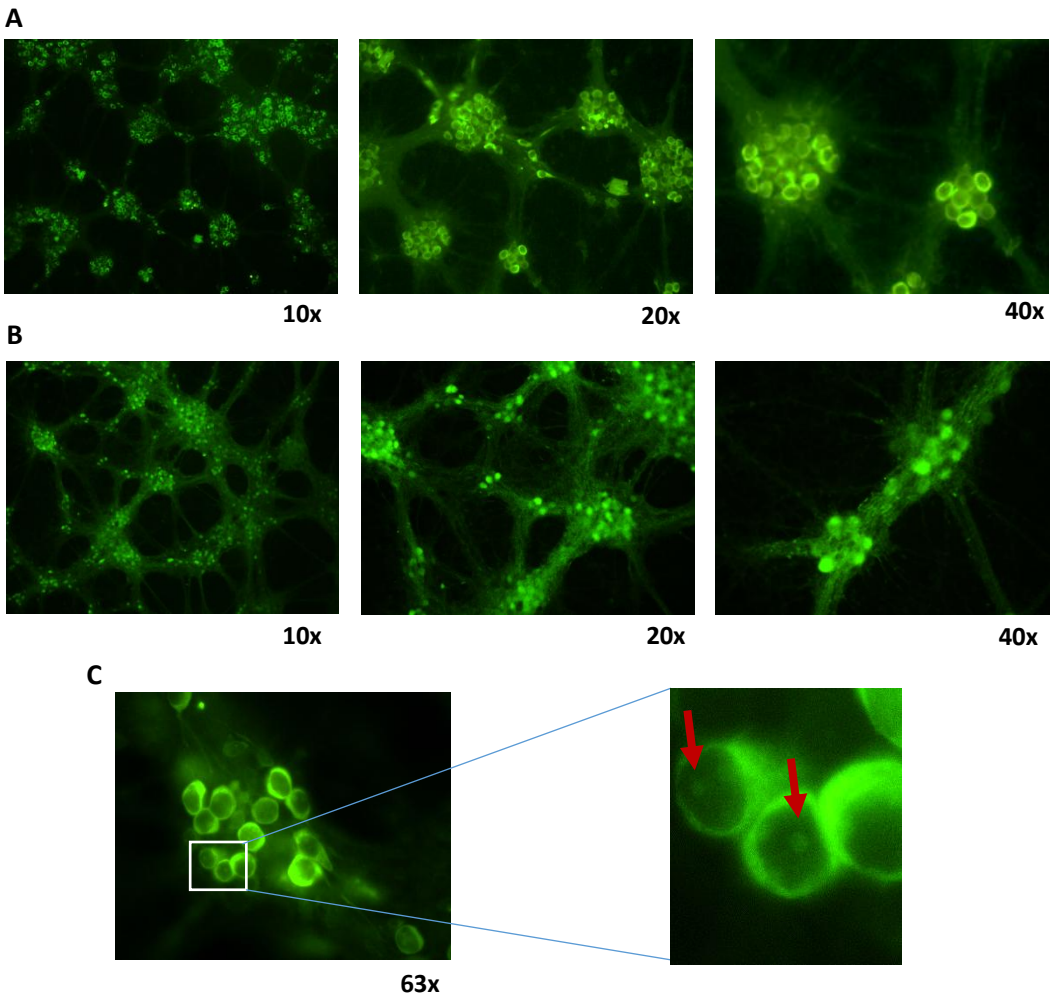


Figure 2.3: EGFP fluorescence in rat hippocampal cultures transduced by lentivirus shown at a range of magnifications. A: CMV-EGFP-L10a; B: CaMKiiα-EGFP; C: CMV-EGFP-L10a with oil immersion and enlarged view of expression within cell bodies. Red arrows indicate areas of enhanced EGFP fluorescence within nuclei.

Figure 2.3 shows the expression of EGFP and its distribution within individual cells following transduction by both viruses. EGFP is expressed in neurons in vitro by both viruses, but with different distributions. EGFP is distributed evenly throughout the cell, whereas EGFP-L10a is most strongly expressed in the soma, excluding the nucleus. However, Figure 2.3C shows a small circular area of concentrated fluorescence within the nucleus, presumed to be the nucleolus. EGFP expression is also observed, though at a much lower level, in the dendrites and possibly elsewhere within the nucleus.

2.3.2.2 *Immunocytochemistry with cell-type markers*

To determine which cell types were targeted by the lentiviruses CMV-EGFP-L10a and CaMKiiα-EGFP, primary hippocampal cultures transduced with the two lentiviruses were incubated with primary antibodies to four cell-type markers (see Table 2.4), which were detected by a red fluorescent-conjugated secondary antibody (CY3). Cells were simultaneously incubated with anti-EGFP and then with a green fluorescent-conjugated secondary (CY2) to boost the EGFP signal.

Table 2.4: Primary antibodies used to stain primary hippocampal cultures transduced by lentivirus.

Marker	Target	Fluorescent Secondary	Dilution	Species	Manufacturer & Product no.
Beta-III tubulin	Neurons	CY3	1:500	Chicken	Millipore AB9354
GFAP	Glia	CY3	1:500	Rabbit	Dako Z0334
CaMKii	Pyramidal neurons	CY3	1:500	Mouse	Millipore 05532
GAD67	Inhibitory neurons	CY3	1:100	Mouse	Abcam ab26116
GFP	Transduced cells	CY2	1:2000	Mouse	Roche 1181446000

Figure 2.4 and Figure 2.5 show, respectively, the co-expression of CMV-EGFP-L10a and CaMKiiα-EGFP with markers of neurons, pyramidal neurons, inhibitory neurons, and glia. Expression of both viruses strongly overlaps with βIII-tubulin expression, but not with GFAP expression. Both viruses also strongly co-expressed EGFP with CaMKiiα. In both cases, EGFP was found in the majority of cells expressing CaMKiiα, but not all cells expressing EGFP expressed CaMKiiα. Few cells strongly expressing GAD67 could be found among the sample transduced by CMV-EGFP-L10a, partially due to high background staining. Nevertheless, there appeared to be little to no overlap with EGFP expression. Among the cells transduced by CaMKiiα-EGFP, small clusters of cells strongly expressing GAD67 were found and these did not co-express EGFP.

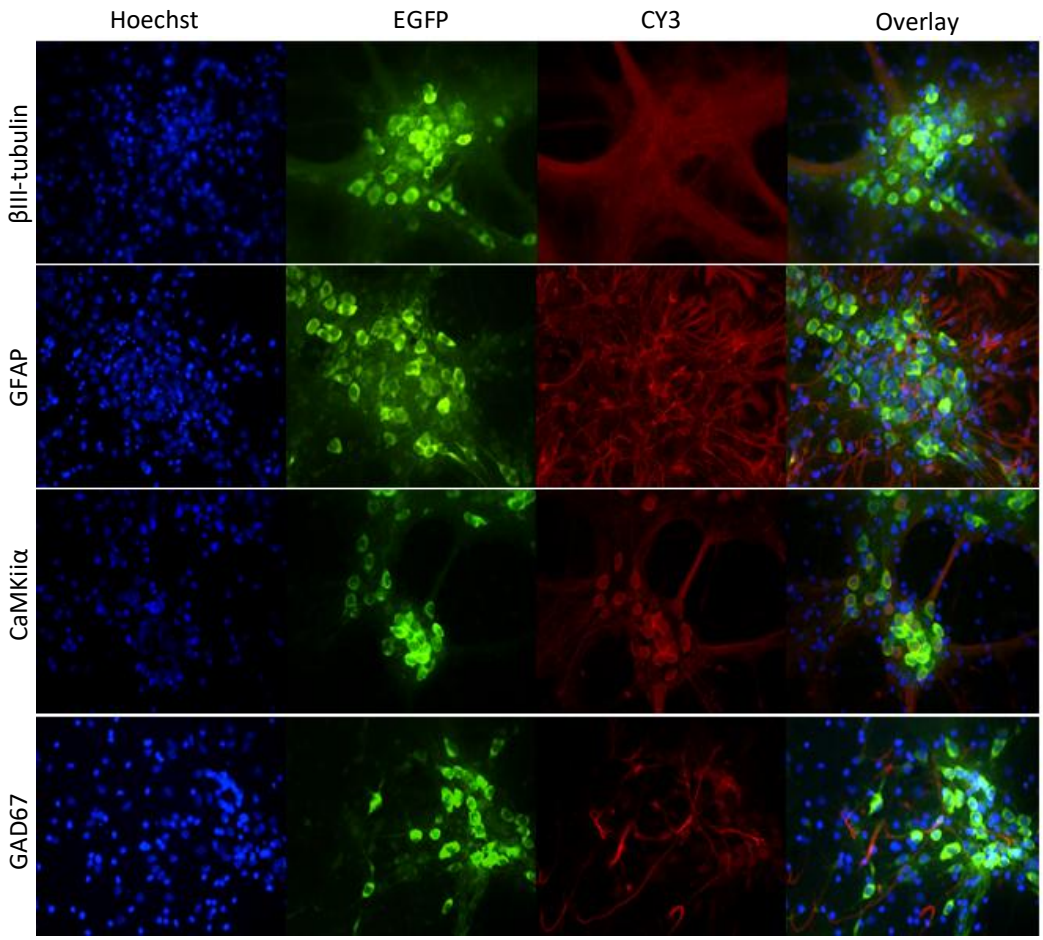


Figure 2.4: Co-expression of cell type markers with lentivirus CMV-EGFP-L10a in hippocampal primary cultures. Cells were incubated with Hoechst (blue fluorescent nuclear stain) and one of four cell type markers visualised with CY3-conjugated secondary antibodies (red): βIII-tubulin (neurons), GFAP (glia), CaMKiiα (excitatory neurons) and GAD67 (inhibitory neurons). Cells stained with βIII-tubulin and GFAP were simultaneously labelled with an antibody to EGFP and a CY2-conjugated secondary (green). Greyscale images were taken at 40x magnification on a fluorescent microscope using three different filters, false colour was applied, and images were overlaid.

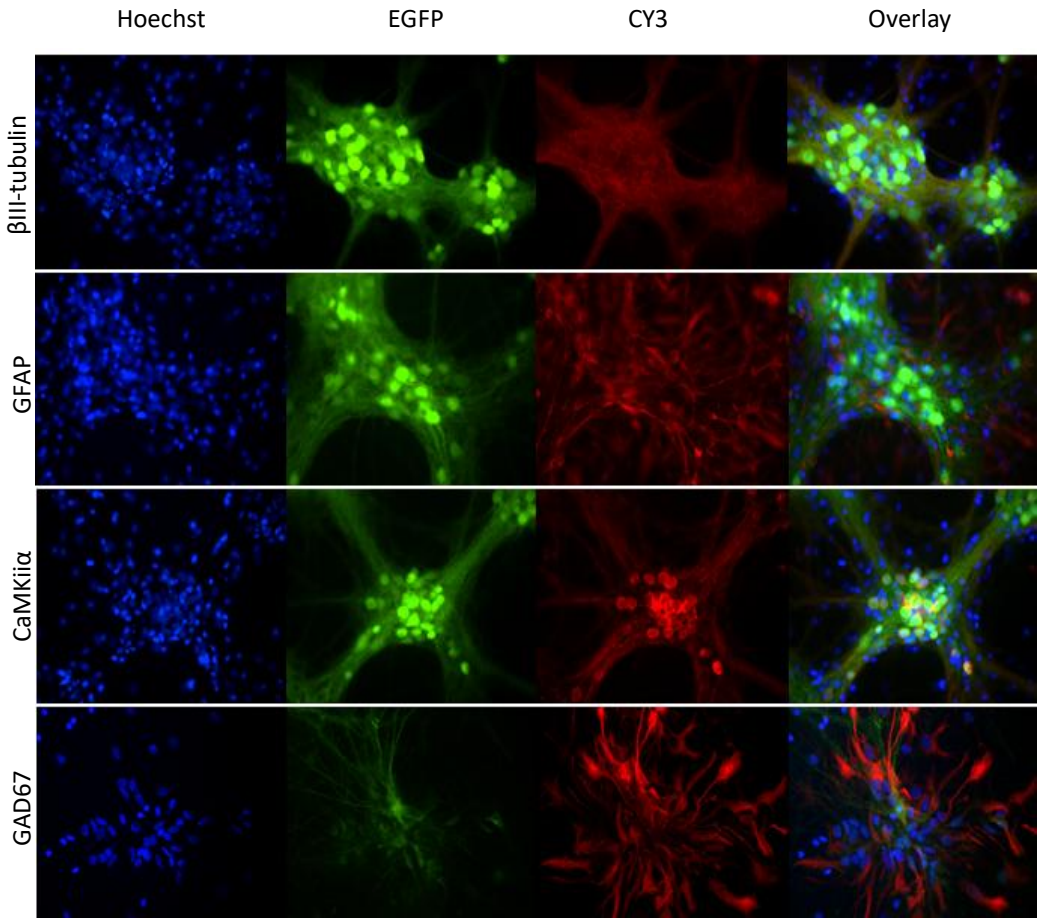


Figure 2.5: Co-expression of cell type markers with lentivirus CaMKiiα-EGFP in hippocampal primary cultures. Cells were incubated with Hoechst (blue fluorescent nuclear stain) and one of four cell type markers visualised with CY3-conjugated secondary antibodies (red): βIII-tubulin (neurons), GFAP (glia), CaMKiiα (excitatory neurons) and GAD67 (inhibitory neurons). Greyscale images were taken at 40x magnification on a fluorescent microscope using 3 different filter cubes, false colour was applied, and images were overlaid.

2.3.3 Optimisation of viral expression in-vivo

2.3.3.1 Viral injections of lentivirus and AAV into CA1 to optimise transduction

To optimise transduction of CA1 in adult Lister Hooded rats, a range of co-ordinates and two viruses were trialled. The anterior-posterior (AP) co-ordinate was -5.3mm posterior to bregma, the latitude (lat.) co-ordinates ranged from ±3.0 to 4.0mm from the midline, and the depth co-ordinates ranged from -3.0 to 4.0mm from dura. The viruses were AAV2/9 (titre 1.11 x 10^{13}/ml) and VSV-G pseudotyped lentivirus with a post-transcriptional regulatory element (WPRE; titre 2.22 x 10^{9}/ml), both expressing EGFP under a CMV promoter.

All subjects (N=8) received 2µl of virus per hemisphere over ten minutes delivered at a range of co-ordinates listed in Table 2.5. In two of the subjects receiving lentivirus, virus delivery was divided into two separate 1µl injections at different sites to determine whether this would improve transduction of CA1. EGFP expression was visualised three weeks post-surgery.

Table 2.5: Co-ordinates of viral injections to optimise targeting of CA1 in rats

			Left hemisphere			**Right hemisphere**		
Subject	**Virus**	**No. of injections**	*AP*	*Lat.*	*Depth*	*AP*	*Lat.*	*Depth*
R1	Lentivirus	2	-5.3	+4.0	-4.0	-5.3	-3.0	-4.0
R2	Lentivirus	4	-5.3	+3.0, +3.5	-3.0, -3.5	-5.3	-3.0, -3.5	-3.0, -3.5
R3	Lentivirus	4	-5.3	+3.5, +4.0	-3.5, -4.0	-5.3	-3.5, -4.0	-3.5, -4.0
R4	Lentivirus	2	-5.3	+3.0	-4.0	-5.3	-4.0	-4.0
R5	AAV	2	-5.3	+3.0	-4.0	-5.3	-4.0	-4.0
R6	AAV	2	-5.3	+3.5	-3.0	-5.3	-3.5	-3.5
R7	AAV	2	-5.3	+4.0	-3.5	-5.3	-4.0	-4.0
R8	AAV	2	-5.3	+3.0	-3.5	-5.3	-3.0	-3.0

Good transduction of the pyramidal cell layer of CA1 was achieved in two of the subjects that received infusions of the AAV virus. Figure 2.6A (rows 1 & 2 respectively) shows a sample of fluorescent images from these subjects, R6 and R8, at 5x magnification. In both subjects, EGFP expression was strong and most of CA1 was transduced (with the possible exception of the area closest to the midline in R8). However, transduction was not confined to CA1 and in subject R8, EGFP fluorescence was brighter in CA3 than in CA1. Figure 2.6B shows images from subject R8 at 10x and 20x magnification, illustrating that EGFP is expressed in pyramidal cell bodies. Few cell bodies outside the pyramidal cell layer were observed to express EGFP, suggesting that transduction of interneurons and glia was minimal.

EGFP was not visibly expressed in CA1 following any of the lentiviral injections. The EGFP signal was very weak in four of the eight hemispheres injected with lentivirus. In the

remaining hemispheres, expression was largely confined to the DG and white matter above CA1, as shown in the images from the left hemisphere of subject R3 in Figure 2.6A, row 3. Overall, both the strength of the EGFP signal and the spread of the virus within the hippocampus were substantially weaker using lentivirus than AAV.

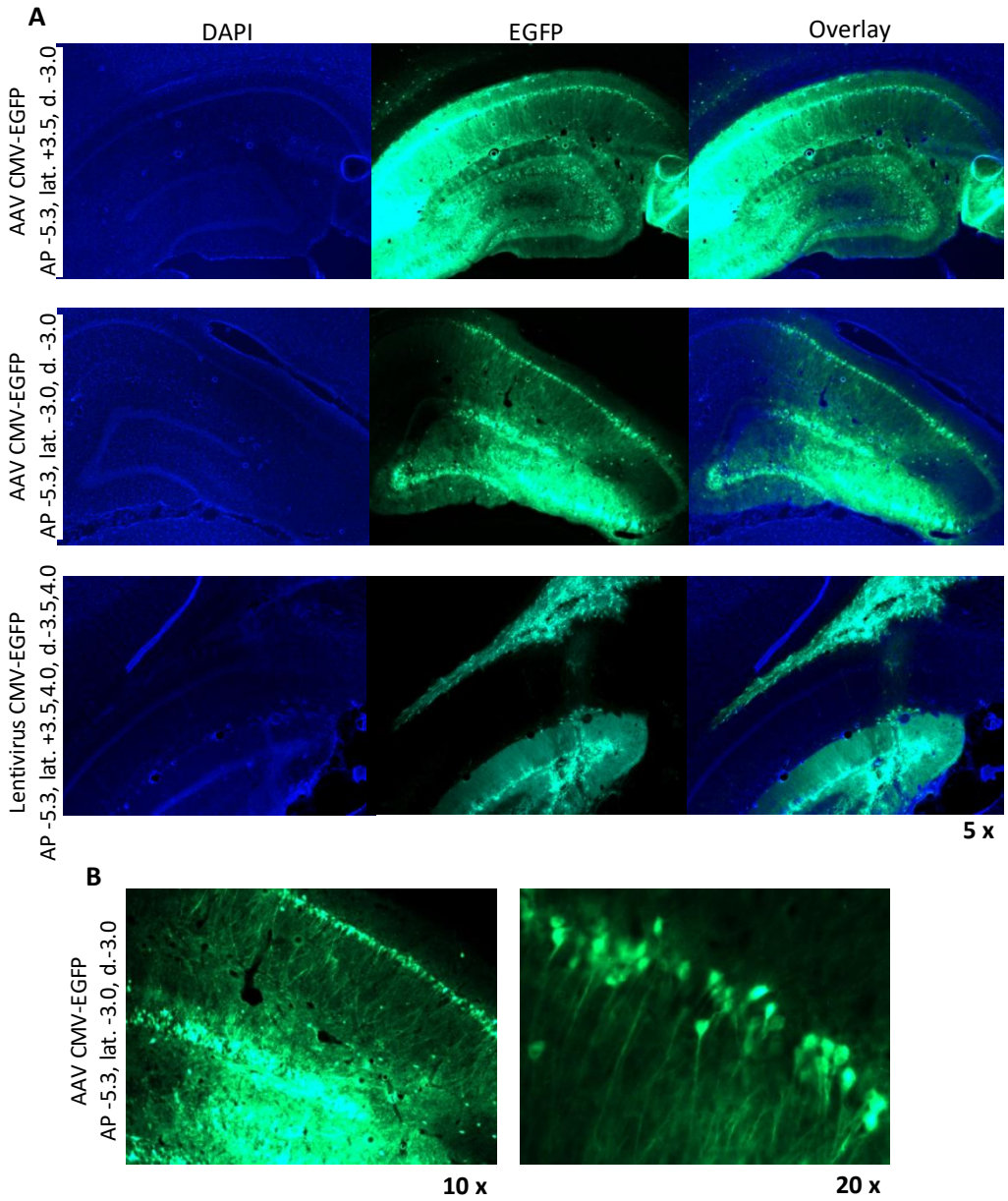


Figure 2.6: Transduction of rat hippocampus by lentivirus and AAV expressing CMV-EGFP injected at varying co-ordinates. A: Fluorescent overlays at 5x magnification show DAPI (blue fluorescent nuclear stain) and EGFP, with red-channel to compensate for background fluorescence. See left-hand side for injection co-ordinates. Rows 1+2: AAV; row 3: lentivirus. B: EGFP expression (no overlay) at 10x and 20x magnification in CA1 transduced by AAV. AP: anterior-posterior co-ordinate; lat.: latitude; d.: depth.

2.3.3.2 Expression of EGFP-L10a in CA1 in-vivo

Injections of lentivirus and AAV9 expressing CMV-EGFP-L10a (titres $3.79x10^8$/ml and $7.76x10^{11}$/ml respectively) were administered to N=4 rats to assess transduction of neurons by the viruses in-vivo and targeting of CA1.

Two sets of co-ordinates were used: those previously optimised (AP -5.3, lat. ±3.5, depth -3.0; 'A') and a further set of co-ordinates intended to better target dorsal rather than ventral or intermediate CA1 (AP -3.3, lat. ±2.0, depth -2.0 from dura; 'B'; from Wong et al, 2005). Lentivirus CaMKiiα-EGFP was also injected into CA1 in a further two subjects to determine whether the CaMKiiα promoter would more strongly transduce the pyramidal cell layer over interneurons and glia compared to the CMV promoter. All subjects received 2µl of virus per region in each hemisphere.

Following infusion of lentivirus CMV-EGFP-L10a, EGFP expression was low in three out of four hemispheres. EGFP expression was observed in the left hemisphere of CA1 using the B co-ordinates, as shown in Figure 2.7, row 1. Transduction was highly specific to the pyramidal cell layer (see Figure 2.8A for higher magnification), but the proportion of CA1 cells transduced was low due to a failure of the virus to spread along either the AP or mediolateral axes.

CA1 was transduced by AAV CMV-EGFP-L10a when the B co-ordinates were used, while the A co-ordinates transduced CA3, as shown in Figure 2.7, rows 2&3 respectively. In both cases the virus transduced the entire lateral extent of the subregion but did not spread into other regions. Within the pyramidal layer, expression was saturated. However, EGFP was also expressed in a small proportion of the cells outside the pyramidal cell layer, as shown in Figure 2.8A&C at higher magnification. Comparison with images from the Allen Mouse Brain Atlas (Lein et al, 2007) of cell type marker gene expression in CA1 measured by RNA in situ hybridisation suggested that these cells may be interneurons rather than glia (Figure 2.9). Transduced cells outside the pyramidal layer appear to be particularly bright, suggesting that they may have been transduced by multiple viral particles. Figure 2.8B shows the spread of virus along the AP axis using the B co-ordinates. AP co-ordinates were estimated from bregma using a rat brain atlas (Paxinos & Watson, 1998). Expression is strongest at the septal (dorsal) pole and progressively weakens from approximately -4.5mm from bregma.

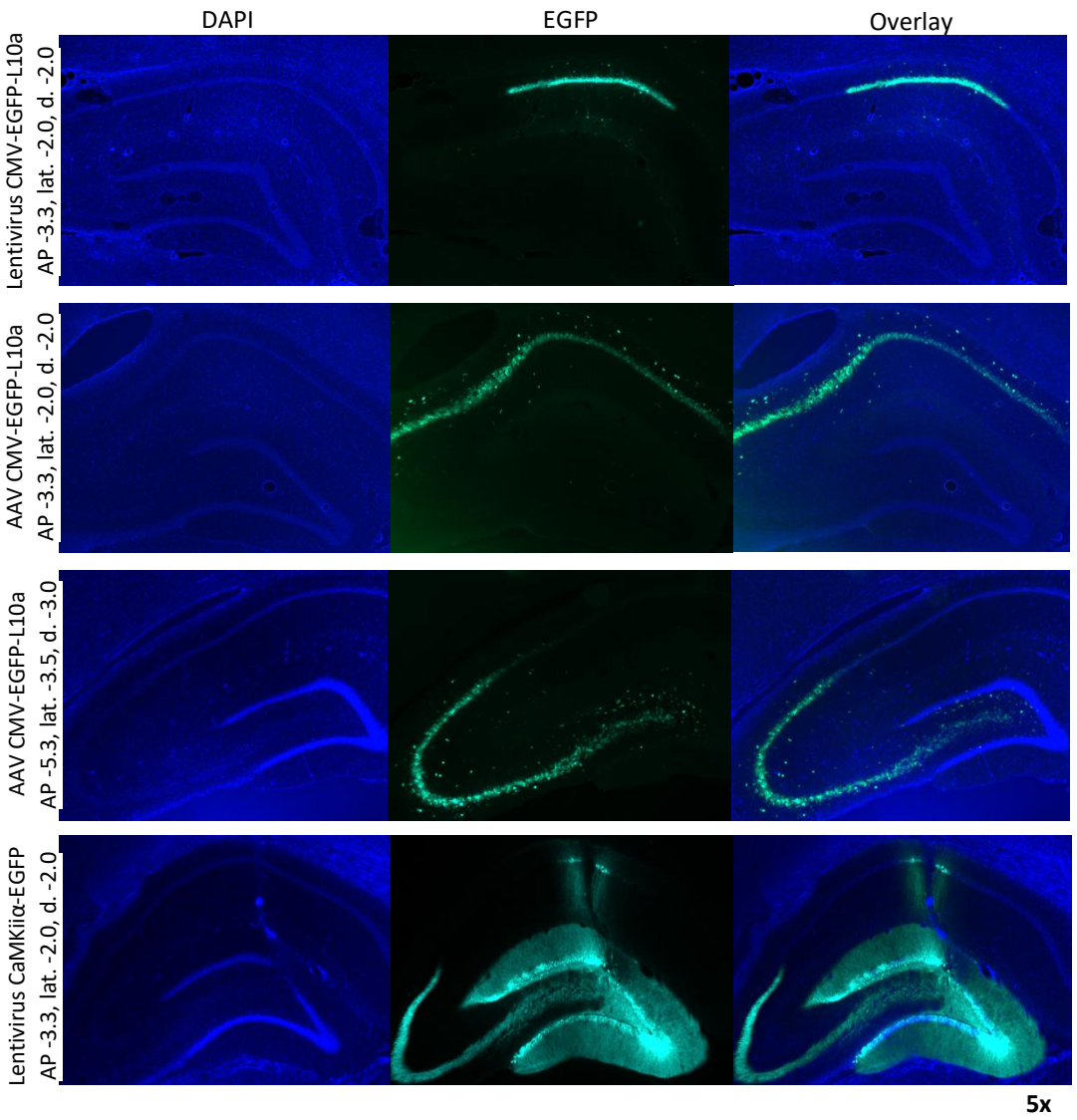


Figure 2.7: Transduction of rat hippocampus by AAV and lentivirus expressing CMV-EGFP-L10a and lentivirus CaMKiiα-EGFP. Co-ordinates were chosen to target CA1. DAPI (blue fluorescent nuclear stain), EGFP, and overlays with correction for background fluorescence are pictured at 5x magnification. AP: anterior-posterior co-ordinate; lat.: latitude; d.: depth.

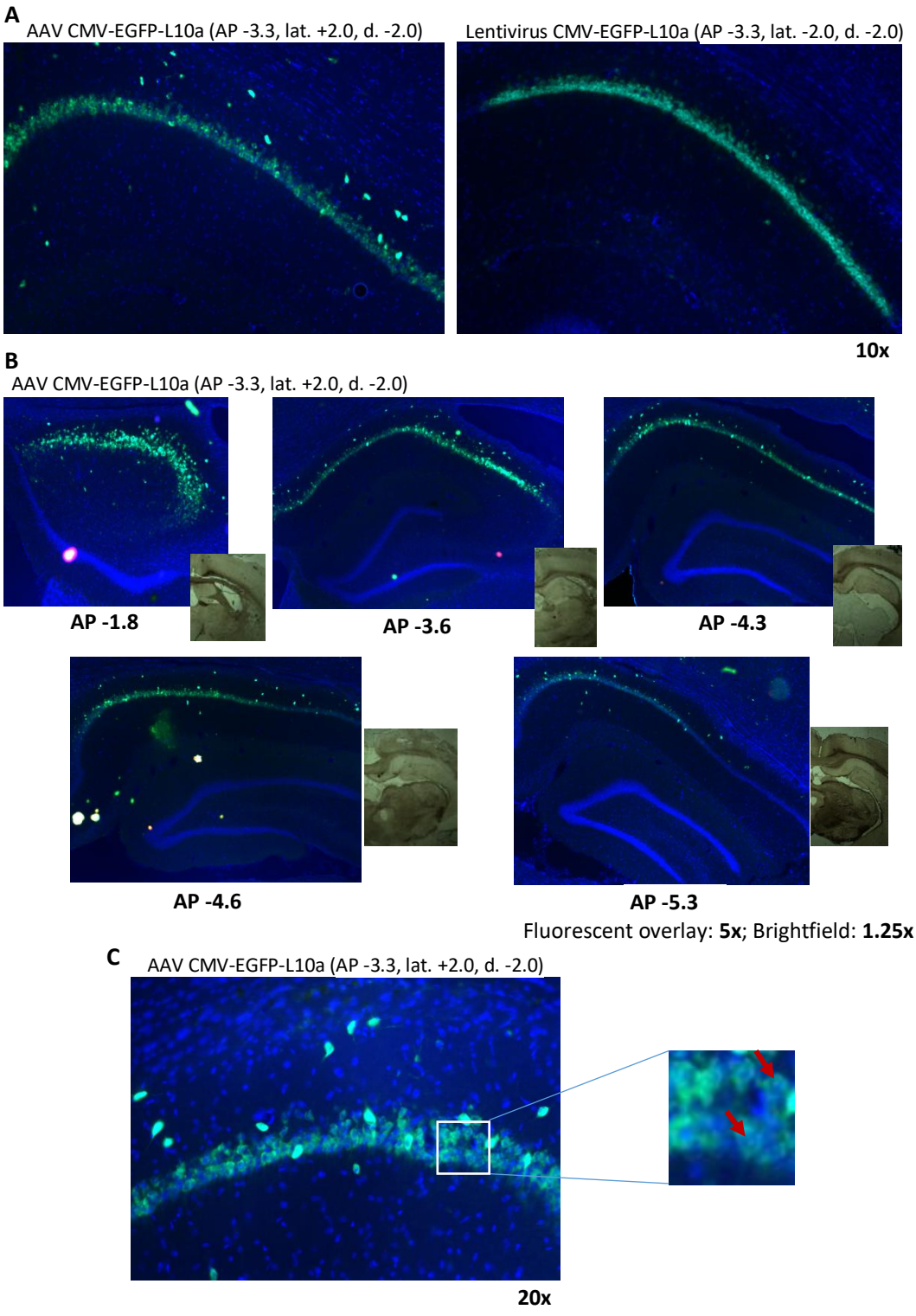


Figure 2.8: Transduction of CA1 by AAV and lentivirus CMV-EGFP-L10a; higher magnification and spread along AP axis. A: Comparison of CA1 transduction by lentivirus and AAV injected at the same co-ordinates, 10x magnification. B: Expression of AAV CMV-EGFP-L10a at a range of AP co-ordinates in one subject to characterise spread. Fluorescent overlays at 5x magnification are presented with estimated AP co-ordinates and corresponding brightfield images at 1.25x magnification. C: CA1 transduced by AAV CMV-EGFP-L10a at 20x magnification, showing distribution of EGFP signal within cell bodies and transduction of cell bodies outside the pyramidal cell layer. Right: Enlarged view of cell bodies with red arrows to indicate presumed nucleoli. AP: anterior-posterior co-ordinate; lat.: latitude; d.: depth.

The lentivirus CaMKiiα-EGFP did not visibly transduce CA1 with either set of co-ordinates, although a small number of neurons were transduced in the CA1 pyramidal layer by the B co-ordinates. Other hippocampal subregions were transduced, as shown in Figure 2.7, row 4.

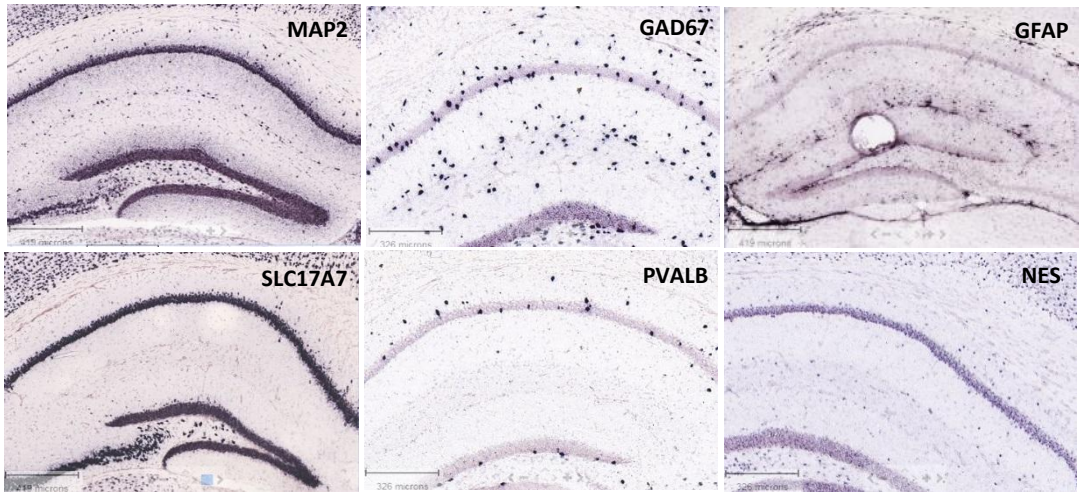
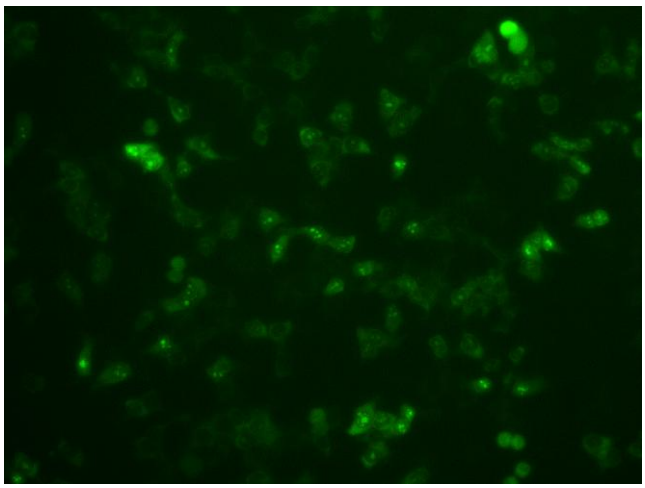


Figure 2.9: Distribution of cell type markers in mouse hippocampus; in-situ hybridisation (ISH) images from Allen Mouse Brain Atlas (Lein et al, 2007). Marker genes are Map2 (neurons), Slc17a7 (excitatory neurons), Gad67 (inhibitory neurons), Pvalb (parvalbumin-expressing interneurons), Gfap (glia) and Nes (glia).

2.3.4 TRAP with virally transduced cells

2.3.4.1 Cultured cells

A stable cell line of HEK293T cells expressing EGFP-L10a was produced and cultured to extract EGFP-tagged polysomes using the TRAP technique. Cultured cells are pictured in Figure 2.10. Approximately 2.5ml of cell lysate was added to 300µl of antibody-conjugated beads. Input and unbound samples of 50µl were removed before and after immunoprecipitation respectively to compare RNA levels.



10x

Figure 2.10: HEK293T cultured cells transduced by lentivirus CMV-EGFP-L10a at 10x magnification

Table 2.6 shows the RNA concentration and absorbance ratios of each sample. The differences in RNA concentration are most likely explained by the differences in sample volumes. After accounting for volume, the concentration of RNA in the input sample prior to nanoprep would have been ~73ng/µl compared to ~11.5ng/µl in the immunoprecipitate (IP). Pure RNA has a 260/280 ratio of ~2.0, suggesting that all samples were free from significant protein contamination. A 260/230 ratio of >2.0 suggests that samples are free from DNA contamination, while a ratio <2.0 suggests genomic DNA contamination and <1.5 suggests other contaminants are present. The unbound sample may therefore contain DNA contamination.

Table 2.6: TRAP RNA from cultured HEK293T cells expressing EGFP-L10a

Sample	Concentration (ng/µl)	260/280 ratio	260/230 ratio
Immunoprecipitate	2898	2.022	2.234
Input	364	2.015	2.020
Unbound	172	2.028	1.800

2.3.4.2 Piloting of TRAP in CA1 tissue

CA1 tissue dissected from Lister Hooded rats injected with AAV CMV-EGFP-L10a bilaterally was used to extract EGFP-tagged polysomes using the TRAP technique. CA1 tissue was pooled across three subjects (six hemispheres) in group 1 and two subjects (four hemispheres) in group 2 to determine the amount of tissue required to generate sufficient RNA for sequencing. Alongside the first group, CA1 tissue was collected from three control subjects not injected with virus. Different methods were used to dissect CA1 for each of the two groups, as discussed in methods. DG and CA3 tissue was also collected from the latter group for analysis. 50µl input and unbound samples were taken to compare with the IP.

Table 2.7 shows the RNA concentration and 260/280 absorbance ratios of each sample measured on a nanophotometer, as well as the RNA integrity number equivalent (RIN^e) measured on an Agilent 2200 TapeStation (Agilent, CA). The viral CA1 IP samples from both two and three rats had an RNA concentration of greater than 10ng/µl, suggesting there is sufficient RNA for sequencing. Counterintuitively, the RNA concentration is greater for the pool of two rats than three, but this is likely due to improvements in technique, such as the use of a more efficient dissection method. When no virus was present, the concentration of RNA was too low to measure, suggesting that levels of background RNA precipitation are low. When tissue from CA3 and DG was used, a high concentration of RNA was found, suggesting spreading of the EGFP-L10a virus outside CA1.

Pure RNA has a 260/280 ratio of approximately 2.0, suggesting that all samples were free from significant protein contamination. RIN^e values ranged from 6.5-7.3, in the intermediate quality range.

Table 2.7: TRAP RNA purified from dissected hippocampi of rats injected with AAV CMV-EGFP-L10a in pilot tests of TRAP method

	Sample	Concentration (ng/µl)	260/280 ratio	RIN[e]
EGFP-L10a	*Immunoprecipitate*	**14.8**	**2.056**	**7.1**
3 animals (6 xCA1)	*Input*	86.6	2.028	7.0
Group 1	*Unbound*	45.5	2.073	7.1
Control	*Immunoprecipitate*	**too low to measure**	-	3.3
3 animals (6 x CA1)	*Input*	101.0	2.048	6.8
Group 1	*Unbound*	72.3	2.057	6.6
EGFP-L10a	*Immunoprecipitate*	**30.3**	**2.054**	**6.9**
2 animals (4 x CA1)	*Input*	29.1	2.028	7.0
Group 2	*Unbound*	22.8	2.036	6.5
EGFP-L10a	*Immunoprecipitate*	97.0	2.059	6.9
2 animals (4 x CA3 & DG)	*Input*	31.5	2.026	7.3
Group 2	*Unbound*	19.6	1.960	6.8

2.4 Discussion

This chapter discusses the cloning of plasmids, the development of AAV and lentiviral vectors expressing EGFP-L10a, and the optimisation of in vitro and in vivo viral expression. The ultimate aim was to use the optimised viral TRAP system to profile the translatome of rat neurons during recognition memory formation.

2.4.1 Virus development

2.4.1.1 Cloning of viral backbone plasmids

Several plasmids were cloned combining EGFP and the TRAP gene, EGFP-L10a, with CMV and CaMKiiα promoters in lentiviral and AAV backbones. Plasmids were produced to be made into viruses expressing either EGFP-L10a to profile the translatome of rat neurons or EGFP as controls.

The first attempt to clone the TRAP gene EGFP-L10a into a lentiviral backbone used PCR to amplify the insert and add 5' and 3' sequences complementary to the digested ends of the vector, followed by a ligation reaction. This method was unsuccessful, which may have been due to a failure of the restriction enzymes to cut close to the end of the polynucleotides. Instead, cloning of the target construct was achieved using a recently developed system, InFusion cloning (Clontech Laboratories, USA), which enables cloning at any site using a novel enzyme that fuses DNA fragments by recognising a 15bp overlap at their ends. This reaction replaces the ligation reaction and removes the need for restriction digestion of the insert and dephosphorylation of the vector.

The success of cloning reactions was confirmed by restriction digestions and electrophoresis and by Sanger sequencing using different primers.

2.4.1.2 Virus production

From the cloned plasmids, three viruses were produced: lentiviruses expressing EGFP-L10a under a CMV promoter and EGFP under a CaMKiiα promoter, and an AAV expressing EGFP-L10a under a CMV promoter.

Lentiviruses were produced in-house and viral titre was measured by qPCR. Titre by fluorescence-activated cell sorting was unsuitable as the CaMKiiα promoter would not induce expression in available cell lines. Estimates of titre based on qPCR were similar across biological and technical replicates, with the exception of one sample that contained a negligible amount of DNA (as measured by nanophotometer).

AAVs were produced and titred by collaborators at King's College London.

2.4.2 Lentivirus expression in-vitro

In-vitro transduction by lentivirus CMV-EGFP-L10a was investigated in rat primary hippocampal cultures. The specificity of the CMV promoter was compared to that of CaMKiiα, using lentivirus expressing EGFP. Both viruses successfully transduced cultured hippocampal cells with a multiplicity of infection of 10 and above.

Fluorescent microscopy revealed that unlike EGFP expression, which was evenly distributed throughout the cell body, nucleus and processes, the expression of EGFP-L10a was primarily concentrated in the soma. Expression was also detectable in neurites and in a concentrated circular region within the nucleus, consistent with the nucleolus. Ribosomes are known to be found in dendrites (Tiedge & Brosius, 1996) and the nucleolus is the site of ribosome production (Cmarko et al, 2008), so expression in this compartment supports the assumption that EGFP-L10a is associated with ribosomal proteins. A similar distribution of EGFP expression was observed in vivo. Moreover, this expression pattern is in line with existing data on EGFP-L10a (Doyle et al, 2008).

Staining of the transduced cultures with a range of cell type specific markers showed that expression of EGFP-L10a under a CMV promoter was primarily confined to βIII-tubulin and CaMKiiα expressing cells, although a small number of individual cells expressing the target gene co-expressed GFAP or GAD67. This was very similar to the expression of EGFP under a CaMKiiα promoter, which co-localised with βIII-tubulin and CaMKiiα, but not with GFAP or GAD67. These co-expression patterns indicate that both lentiviruses primarily transduced excitatory pyramidal neurons over interneurons and glia, with the CaMKiiα promoter having greater specificity. However, assessment of co-expression of the transgenes with GAD67 was impaired by high background staining of cells by the GAD67 antibody and the low number of interneurons present.

Lentivirus was also used to produce a stable HEK293T cell line expressing EGFP-L10a due to its ability to integrate into the genome.

Overall, these results demonstrate that lentivirus CMV-EGFP-L10a expresses the target construct. Only slight differences were observed in the pattern of expression across different cell types using the CaMKiiα compared to the CMV promoter in lentivirus, suggesting that the CMV promoter is suitable for the present application. However, the cell type specificity of the two promoters may differ with AAV. AAV was not used for transduction of primary cultures due to a potential contamination of the virus with mycoplasma during its production. Moreover, expression of AAV in cultured neurons is not predictive of in vivo results (Carter, 2004). Expression patterns may also differ in-vivo in adult rodents with lentivirus (Smith et al, 2000) and so analysis of the results of viral injections is essential.

2.4.3 Optimisation of viral expression in-vivo

Viral injections were administered to adult rats to optimise co-ordinates, viral vector (AAV or lentivirus), and promoter for the transduction of CA1 pyramidal neurons.

2.4.3.1 Optimisation of injection co-ordinates

A range of co-ordinates were tested to optimise transduction of the CA1 pyramidal layer using both AAV and lentiviruses expressing EGFP or EGFP-L10a.

Initially, a range of co-ordinates with the same AP but varying latitude and depth were tested to optimise transduction of CA1 using AAV and lentivirus expressing EGFP under a CMV promoter. However, the results were inconsistent. Strong EGFP expression in the CA1 pyramidal layer was obtained using AAV with two sets of co-ordinates which had

the same AP and depth and similar latitude co-ordinates, but subsequent injections using the same co-ordinates with AAV CMV-EGFP-L10a transduced CA3, suggesting that these co-ordinates are not robust to measurement error and minor variations in physiology (e.g. the location of bregma). CA1 neurons were also not transduced by lentivirus using similar co-ordinates and transduction was not improved by injecting at two sites per hemisphere with the same overall volume of virus. It is also worth noting that differences in expression between viruses expressing EGFP and EGFP-L10a were affected by the differing distributions of the transgenes within cells, the higher titres of the EGFP viruses used, and for AAV, the difference in serotype. The serotype of the EGFP virus was AAV2/9, meaning it had the ITRs of AAV2 with the capsid of AAV9, whereas the EGFP-L10a virus was AAV9.

An alternative set of co-ordinates was trialled in subsequent injections. These more anterior co-ordinates from Wong et al (2005) were optimised for the transduction of dorsal CA1 as spatial tasks such as associative recognition primarily recruit dorsal rather than ventral hippocampus (Fanselow & Dong, 2010). Using these co-ordinates, the CA1 pyramidal layer was transduced by AAV CMV-EGFP-L10a in both hemispheres injected and by lentivirus CMV-EGFP-L10a in one of two hemispheres injected, although only a small number of cells in CA1 were transduced by lentivirus CaMKiiα-EGFP using these co-ordinates. Overall, this set of co-ordinates was determined to be more reliable for the transduction of CA1.

2.4.3.2 Comparison of virus type and promoter

Injections of lentivirus and AAV expressing EGFP or EGFP-L10a under a CMV promoter were compared to determine which type of virus would produce the strongest and most specific expression in CA1 pyramidal neurons. Additional injections of lentivirus expressing EGFP under a CaMKiiα promoter were made to compare the cell type specificity of promoter activity as CaMKiiα expression is generally restricted to pyramidal neurons (Dittgen et al, 2004) but may drive weaker expression compared to the constitutive promoter CMV (Kuroda et al, 2008).

Good transduction of CA1 was obtained with lentivirus CMV-EGFP-L10a in one case. Expression was strong and highly specific to the pyramidal cell layer. However, spread was insufficient along both the AP and mediolateral axes. Transgene expression was weak following several of the injections of lentivirus with a CMV promoter. Expression of lentivirus with a CaMKiiα promoter was strong in comparison with injections of EGFP with a CMV promoter and appeared to only be expressed in pyramidal neurons in CA1, but very few CA1 cells were transduced. Overall, the results of lentiviral injections were inconsistent.

AAV expression in contrast was very strong. This is at least in part due to the much higher titres of the AAVs compared to the lentivirus as AAV particles are much smaller. Additionally, AAV9 can be both retrogradely and anterogradely transported (Castle et al, 2014) and so spreads further than lentivirus. The titre of AAV 2/9 CMV-EGFP was particularly high and this virus spread across hippocampal subregions. For the present application it would be preferable to transduce only CA1 to avoid contamination from

other regions, although non-specific expression can be compensated for to some extent by careful dissection of CA1.

AAV9 CMV-EGFP-L10a had a lower titre and expression was confined to a single subregion (CA3 or CA1) at all four hippocampal injection sites. Promisingly, AAV transduced the entire lateral extent of both subregions, and spread well through dorsal CA1 in the posterior direction. However, cells outside the pyramidal layer were also transduced, most likely interneurons rather than glia based on their distribution, but this could be confirmed with IHC using markers of interneurons and glia. Targeting may be improved using a CaMKii promoter.

2.4.3.3 Summary

AAV appears to give both specificity and spread appropriate for the transduction of CA1 and a set of reliable co-ordinates was found. Transduction of CA1 by lentivirus was more specific when successful, but transduction was unreliable and the virus did not spread very far from the injection site. Maximising the number of cells transduced is necessary to obtain enough IP RNA for sequencing, target more of CA1, and minimise the number of animals needed. AAV has also been used by other research groups for delivery of the TRAP gene (Kratz et al, 2014; Nectow et al, 2017).

Specificity of AAV expression may be improved if necessary by using a different promoter such as CaMKiiα, though the number of cells transduced outside the pyramidal layer was low with CMV due to the tropism of the virus. In the subsequent experiment, virus expression was checked at periodic intervals to ensure continued expression of the virus and good transduction of CA1.

2.4.4 Piloting of TRAP and dissection

2.4.4.1 Piloting of TRAP in cultured cell lines

Ribosome-bound mRNA was extracted from cultured HEK293T cells expressing EGFP-L10a using the TRAP method (Doyle et al, 2008). The concentration of RNA obtained was very high, which can be explained by the large number of cells used. Contamination from DNA, protein and other sources indicated by absorbance ratios was low.

2.4.4.2 Piloting of TRAP in tissue

The TRAP method was used to extract polysomes from CA1 tissue dissected from rats injected with AAV CMV-EGFP-L10a and from control animals that had not received injections. The aim was to determine the number of pooled CA1 samples needed to generate a sufficiently large yield of RNA and the background level of RNA purification.

A sufficiently large yield of RNA for sequencing, considered to be a minimum of 10ng/µl, was purified from CA1 tissue pooled across both three animals (six hemispheres) and two animals (four hemispheres). A greater yield of RNA was obtained with two animals than with three, likely due to refinements in technique, in particular a faster dissection method. When tissue that had not been transduced with virus was used, the RNA concentration was too low to measure, suggestive of low background precipitation of RNA. However, RNA was precipitated from DG and CA3 in animals injected with AAV

CMV-EGFP-L10a targeting CA1 suggesting spread of the virus outside the CA1 subregion and confirming the necessity of a subregion-specific dissection.

2.4.5 Conclusions

Lentivirus and AAV viruses were developed to transduce rat hippocampal neurons. The AAV virus with a CMV promoter was determined to transduce more cells and have a more reproducible expression pattern in vivo compared to the lentiviruses tested. Lentivirus expressing CMV-EGFP-L10a was found to be suitable for the transduction of pyramidal neurons in rat primary hippocampal cultures.

Injection conditions were optimised for selective transduction of dCA1 neurons in rats and suitable co-ordinates were obtained through a range of pilot injections.

In a pilot test of the TRAP method, RNA was immunoprecipitated from rats injected with AAV CMV-EGFP-L10a at the chosen co-ordinates and not from control rats that were not injected with virus, suggesting that extracted RNA was bound to EGFP-tagged ribosomes. A sufficient quantity of RNA for sequencing was obtained by pooling CA1 tissue from two rats. RNA was also extracted from DG and CA3, indicating that a subregion-specific dissection is necessary. In addition, highly concentrated RNA was obtained from a stable HEK293T cell line transduced by lentivirus using the TRAP method.

Chapter 3 : A method of translatome profiling during associative recognition memory formation

3.1 Introduction

This chapter presents the experiments that aimed to refine a method to profile the translatome of CA1 neurons during the formation of associative recognition memory in rodents.

3.1.1 RNA sample preparation

To investigate gene expression, several steps are required to prepare purified RNA for sequencing. The following sections outline these steps.

3.1.1.1 RNA quality assessment

RNase enzymes are abundant in the environment and their action must be prevented to maintain RNA integrity and sample quality. RNA quality has traditionally been estimated using the ratio of the quantity of intact 28S rRNA to 18S rRNA measured by gel electrophoresis, based on the assumption that higher integrity samples will contain a higher proportion of 28S rRNA as this species is degraded more rapidly (Auer et al, 2003). A more accurate and reliable method of assessing RNA integrity was developed using the Agilent Bioanalyser 2100 (Agilent, CA, USA), which detects fluorescence intensity of samples separated by electrophoresis in microfluidic chambers. Schroeder et al (2006) used a machine learning approach to reproduce expert ratings of RNA integrity from electrophoretic profiles and assign a user-independent RNA integrity number (RIN) between 1 and 10 based on several factors including the total RNA ratio, the 28S:18S ratio, and the 28S peak height. In the present study, a RIN threshold was determined to ensure that input RNA was of suitable quality for the application, as recommended by Agilent (Mueller et al, 2016).

3.1.1.2 Amplification and ribosomal depletion

Where limited genetic material for sequencing is available, an additional amplification step can be performed prior to adapter ligation and library amplification.

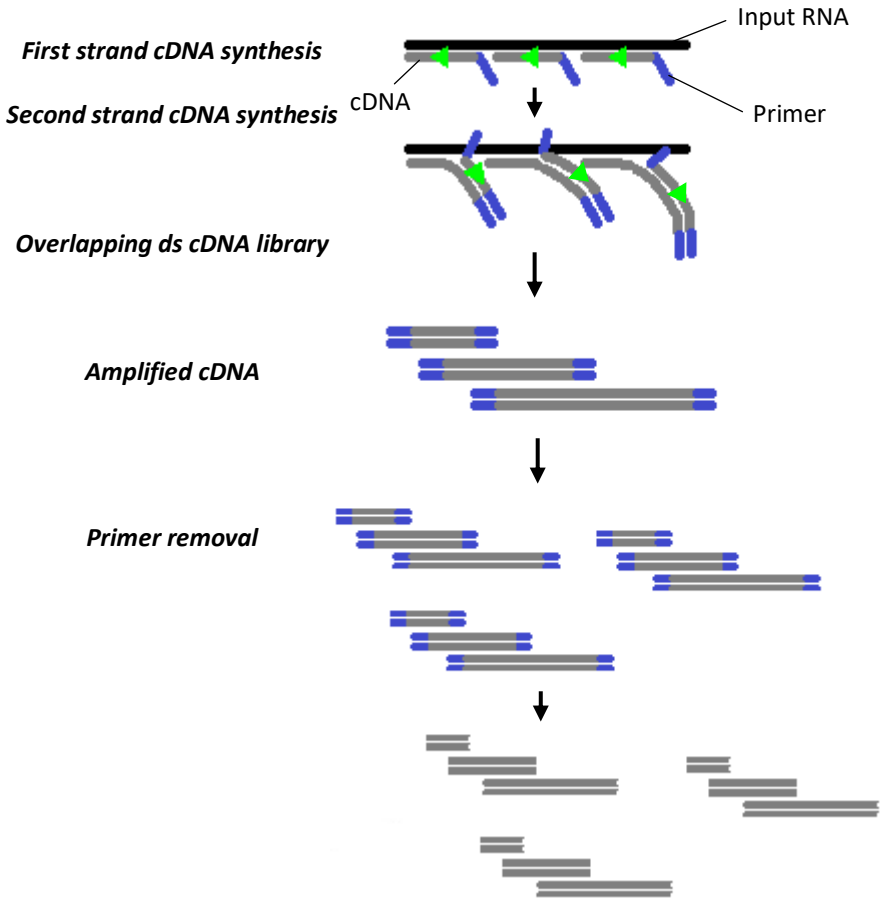


Figure 3.1: Schematic of SeqPlex RNA Amplification kit procedure. Primers with a degenerate 3' end and defined universal 5' end are used to synthesise a double stranded cDNA library and amplify the whole transcriptome. Primers are removed prior to library preparation.

The whole transcriptome amplification (WTA) method uses pseudo-random hexamer (6nt) primers to perform relatively unbiased amplification of all sequences (Langmore, 2002). As ribosomal RNA (rRNA) accounts for approximately 90% of total RNA, an additional rRNA depletion step may be performed following amplification (O'Neil et al, 2013). Alternatively, mRNA can be selectively amplified using primers that bind to the poly-A tail ligated to the 3' end of all mRNAs prior to translation (Shanker et al, 2015). Two methods that amplify both species but amplify mRNA more effectively than rRNA are the SeqPlex RNA Amplification kit (Sigma-Aldrich, St. Louis, MO, USA) and the Ovation RNA-Seq System V2 (NuGEN Technologies Incorporated, San Carlos, CA, USA). SeqPlex is a WTA method using primers that have a universal 6nt sequence at the 5' end and a semi-degenerate 3' end and are removed after amplification, as illustrated in Figure 3.1. Ovation uses proprietary DNA/RNA chimera primers designed to reduce amplification of rRNA in combination with single primer isothermal amplification and produces single-stranded cDNA that is converted to double-stranded cDNA by random hexamer priming. A recent study evaluating these amplification methods against two methods based on poly-A selection found that poly-A selection lead to more complete removal of rRNA but may bias expression data (Shanker et al, 2015). Poly-A selection is also ineffective for amplifying degraded samples as only the 3' ends of mRNA transcripts are bound (Adiconis et al, 2013). A WTA-based method was therefore used in this thesis

and the amplification protocol was optimised to ensure retention of differential expression patterns.

3.1.2 Data processing

Raw RNAseq data can be transformed into a matrix of gene expression counts by first mapping reads to a reference genome or transcriptome and then counting the number of reads aligned to each feature of interest, usually known genes. There are many alternative pipelines for RNAseq data analysis and procedures are not standardised (Conesa et al, 2016). This section discusses methods for processing RNAseq data.

3.1.2.1 Quality assessment of sequenced libraries

Quality assessment and removal of low-quality sequences are important steps prior to mapping. Quality scores assigned by the base-calling algorithm, the proportion of undetermined bases (N content), and the distribution of sequence lengths in a sample (which should be uniform on the Illumina platform) are the main indicators of sequencing quality but do not indicate the quality of input libraries.

Factors of bias in input libraries include adapter contamination, poor library diversity, inconsistent or excessive amplification, and guanine-cytosine (GC) bias. In an unbiased library, the proportion of each nucleotide should be consistent across all read positions and close to 25% each, depending on the genome GC content. Positional biases towards short sequences (k-mers) may indicate adapter or primer contamination. Overall, the per-sequence GC content should be normally distributed. These analyses can be performed by FastQC (Andrews, 2010), a popular quality control application for NGS data. FastQC additionally returns overrepresented sequences accounting for more than 0.1% of reads, which may indicate contamination or a library with low diversity, potentially due to over-amplification.

Another indicator of library quality is the level of sequence duplication. FastQC v0.11.5 returns the proportion of all sequences and distinct sequences at varying levels of duplication. In RNAseq, some sequence duplication is expected because expression levels differ drastically between genes and high-count genes must be sampled many times to generate sufficient power to detect differential expression and low-count genes. However, the duplication profile is also affected by library diversity, amplification, and the presence of contaminants. To distinguish between duplication from biological and technical sources, the relationship between duplication rate and read density can be examined using the dupRadar package (Sayols et al, 2016). For a library with mainly biological duplication, the duplication rate will be low for genes with low read density and increase proportionally. For samples with extensive technical duplication on the other hand, the duplication rate will be initially high and rise little with increasing read density. A logistic regression model can be fitted to the data, such that the intercept indicates the duplication rate for low-count genes and the slope indicates the progression of duplication rate with read density. FastQC and dupRadar were used for quality assessment of the present dataset.

3.1.2.1.1 Alignment

The alignment of RNAseq data is a challenging task because individual genes can have multiple alternative transcripts and splice junctions are often not known in advance, leading to ambiguities in read mapping (Engström et al, 2013). There are many different tools for spliced alignment, the most widely used of which is Tophat2 (Kim et al, 2013), based on the ungapped aligner Bowtie2 (Langmead & Sazberg, 2012). Other spliced aligners include STAR (Dobin et al, 2013) and GSNAP (Wu & Nacu, 2010).

Systematic comparisons of splice-aware alignment tools have shown that Tophat2 achieves relatively poor results compared to other tools, both in terms of speed and accuracy (Hayer et al, 2015; Baruzzo et al, 2016). This is in part because Tophat2 does not truncate reads for partial alignment and therefore has low tolerance of mismatches (Engström et al, 2013). The performance of Tophat2 can be improved by using Bowtie2 to perform local alignment of unaligned reads or by pooling all available samples to find novel splice sites and using this data as input for re-mapping (Engström et al, 2013). However, similar or better performance can be achieved by other tools such as STAR and GSNAP without these time-consuming additional steps (Hayer et al, 2015; Baruzzo et al, 2016). STAR is an extremely fast algorithm and achieves consistent results with default options but requires at least 30GB of random-access memory to run (Engström et al, 2013; Baruzzo et al, 2016).

Alternatively, sequences can be assembled directly from RNAseq data, such as by the Trinity pipeline (Grabherr et al, 2011), which is divided into three modules outlined in Figure 3.2 (adapted from Haas et al, 2014). Inchworm performs the initial assembly of overlapping reads into contigs; Chrysalis clusters overlapping contigs and determines the divergent and shared portions of these sequences by constructing de Bruijn graphs (see Figure 3.2), and Butterfly constructs full-length transcripts representing genes and alternative isoforms. Trinity was used in the present study to reconstruct contaminant sequences.

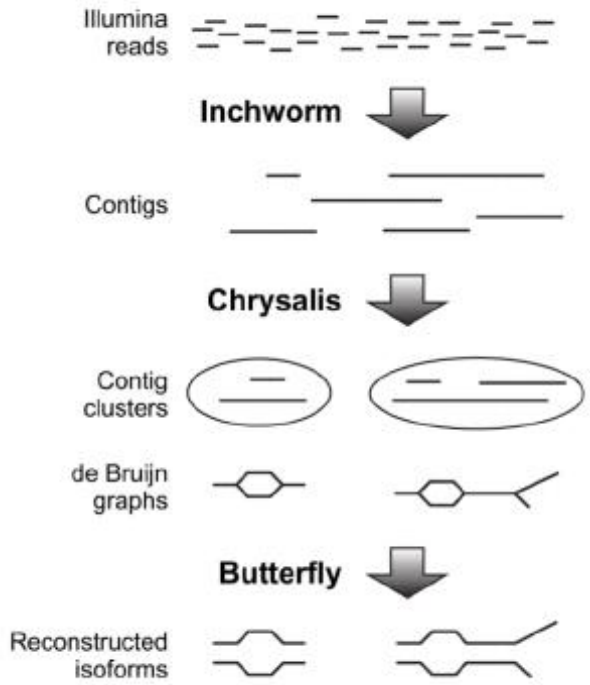


Figure 3.2: Schematic of the Trinity transcriptome assembly pipeline from Haas et al (2014). The three modules of the Trinity pipeline, Inchworm, Chrysalis, and Butterfly, assemble overlapping reads into contigs, cluster overlapping contigs and construct de Bruijn graphs representing the divergent and shared portions of overlapping sequences, and reconstruct gene isoforms from paths through the de Bruijn graphs.

3.1.2.1.2 Quantification of gene expression

Tools for the quantification of aligned reads include Cufflinks (Trapnell et al, 2010) and HTseq (Anders et al, 2014). HTseq simply counts the uniquely aligned reads at each gene locus specified by a genome annotation model, whereas Cufflinks measures gene expression in units of reads per kilobase per million mapped reads (RPKM; Mortazavi et al, 2008) and is suitable for the analysis of alternative splicing. RPKM is a continuous measure that accounts for gene length as longer genes must produce more reads to express the same amount of protein. However, differential expression analysis is biased by correcting for gene length (Oshlack and Wakefield, 2009) and popular downstream analysis tools, such as edgeR (Robinson et al, 2010) and DEseq (Anders & Huber, 2010), use discrete probability distributions and thus require raw counts as input. HTseq was therefore used for gene expression quantification in this chapter and units of counts per million (CPM) were used to adjust for the total number of reads in each library to compare expression of cell type markers.

3.1.3 Present work

This chapter presents the design of an experiment to profile the translatome of rat CA1 neurons during associative recognition memory formation using a paired viewing procedure combined with a viral method of TRAP, the development of which is described in Chapter 2. Here, the pipeline from behavioural methods and pilot studies,

to RNA sequencing and quality assessment, to the generation of gene expression count matrices is presented.

The paired viewing procedure was chosen as the method by which to present series of novel and familiar configurations of images to rats to investigate differences in gene expression profiles up to three hours following stimulus presentation. This procedure was adopted to maximise control over the effects of extraneous variables on gene expression, as discussed in section 1.1.3.2.1. As the paired viewing procedure does not include a behavioural measure of memory performance, rats were also tested using a spontaneous test of hippocampal-dependent recognition memory, the object-in-place test.

TRAP samples were sequenced on the Illumina platform, following RNA amplification to ensure availability of sufficient material in all samples for quality assessment, library preparation, and qPCR. A RIN threshold was determined based on quality assessment of RNAseq data. The analysis pipeline was optimised to generate gene expression count tables for downstream analysis and the enrichment of transcripts from CA1 neurons in TRAP immunoprecipitate samples was confirmed.

3.2 Methods

3.2.1 Subjects

Adult male Lister Hooded rats (Harlan, UK) aged 3-4 months and weighing at least 300g were used for all experiments. Animals were maintained on a 12-hour reverse light cycle with experiments performed during the active (dark) phase. All procedures were performed under licence in accordance with the UK Animals (Scientific Procedures) Act 1986.

3.2.2 Experimental Design

Figure 3.3 shows a schematic of the experiment. Briefly, rats were stereotactically injected with an AAV vector expressing CMV-EGFP-L10a into area CA1 of the hippocampus bilaterally (using the co-ordinates described in 3.2.3 below) and trained in the paired viewing procedure. On the test day, the final day on which images were shown, the subjects were presented either with novel images on one side and familiar images on the other side or, in the case of no-image control subjects, with a black screen on both sides. CA1 tissue was dissected 10, 30, 90, or 180 minutes after the end of the pairing viewing. Tissue from four hemispheres was pooled per sample for tissue homogenisation and immunoprecipitation of EGFP-tagged polysomes.

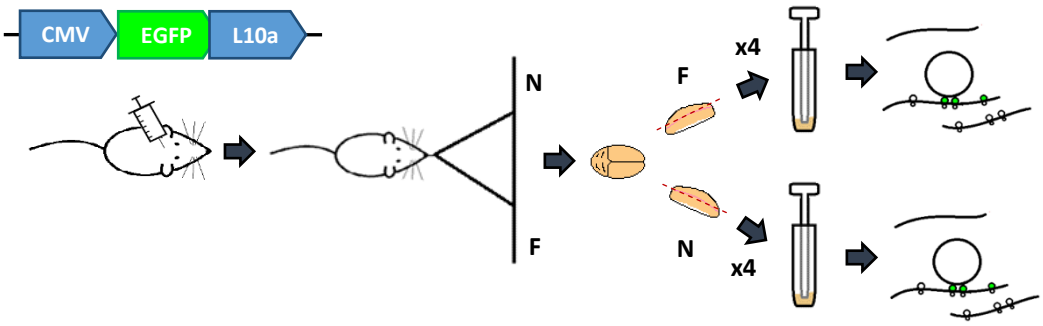


Figure 3.3: Schematic of experimental design. Subjects first received intracerebral injections of virus expressing CMV-EGFP-L10a and were then trained in the paired viewing procedure. CA1 was dissected from both hemispheres after subjects viewed novel (N) and familiar (F) stimuli. Tissue pooled from four subjects was homogenised and RNA bound to EGFP-tagged polysomes was immunoprecipitated using magnetic beads coated in anti-EGFP.

3.2.3 Viral injections

Bilateral injections of AAV9 expressing CMV-EGFP-L10a were made into dCA1 at the co-ordinates AP -3.3mm from bregma, latitude ±-2.0mm from the midline, and depth -2.0mm from dura. (See section 2.2.3 for details of surgical methods.)

3.2.4 Paired viewing

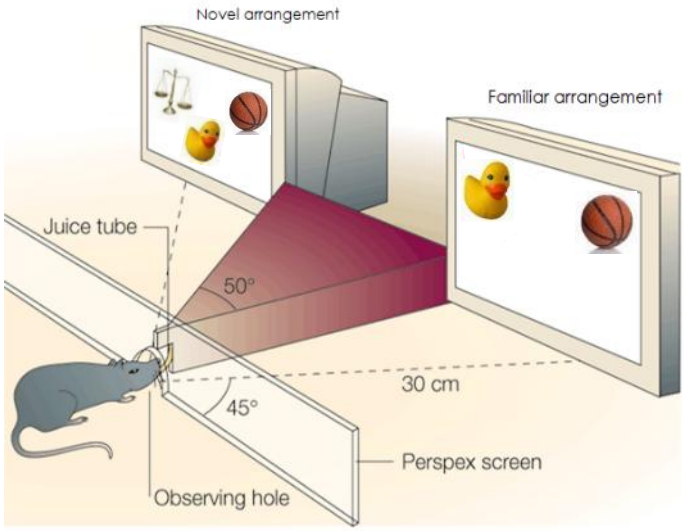


Figure 3.4: Schematic of paired viewing setup with an example pair of novel and familiar arrangements. Figure adapted from Brown & Aggleton (2001).

3.2.4.1 Apparatus

For the pairing viewing procedure, rats were placed in a 30x30x35cm plexiglass box which comprised three black walls and one transparent wall and had a metal grid covering the top. The transparent wall had a centrally-located observing hole (diameter 3cm; height 6cm from floor) and faced two screens at a distance of 30cm. An infrared beam ran across the observing hole and was broken by head entry. As shown in Figure 3.4 and Figure 3.5, a black triangular divider was placed between the two screens such that each screen was visible to only one eye when the subject was in position. A nozzle 7cm from the bottom of the divider delivered individual drops of blackcurrant juice upon the opening and closing of a valve. The valve, infrared beam and detector, and the monitors were connected to a Raspberry Pi computer inside a control box and controlled by scripts written in Python 2.7.

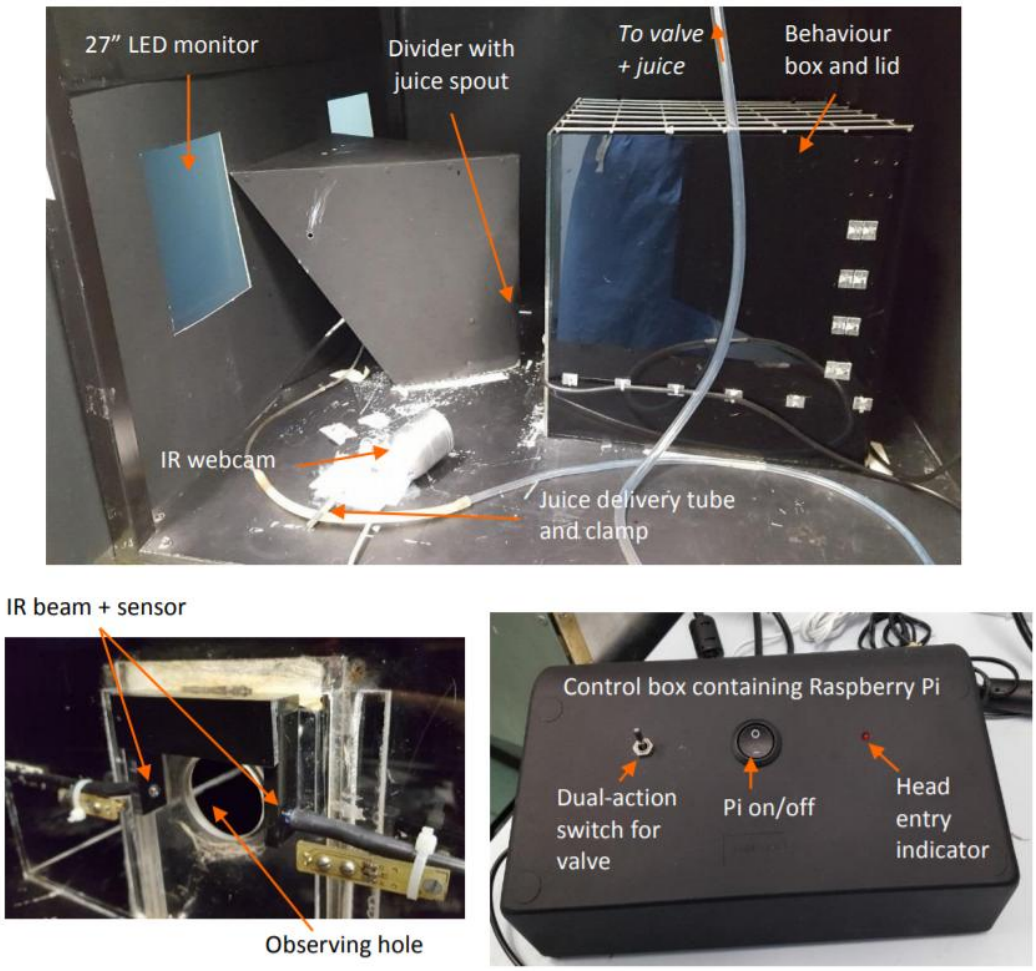


Figure 3.5: Labelled photographs of paired viewing setup. i) Behavioural box and screen for image presentation. ii) Observation window. iii) Control box containing Raspberry Pi connected to monitors, IR beam and detector, and valve controlling juice delivery.

3.2.4.2 Stimuli

Image lists were constructed from a total of 54 images obtained from Clipart. The images were divided into three sets. Each set contained six sub-sets of three images that were always presented together to comprise an image arrangement (see Figure 3.6 for examples). The image arrangement was made by positioning three items on a 9x9 grid. Lists for each training session consisted of five different configurations of each sub-set of three images in a set, with component images always in the same order from left/top to bottom/right. The same images were used for two or three different lists and there were seven lists in total (designated "A" to "F" and "K"). In lists using the same set of images, the positions of two items were exchanged while the third remained in place. Lists A and K were used for the novel and familiar conditions and contained the same component images. Four example images from lists A and K are shown in Figure 3.6. Lists B to F contained images from the other two sets.

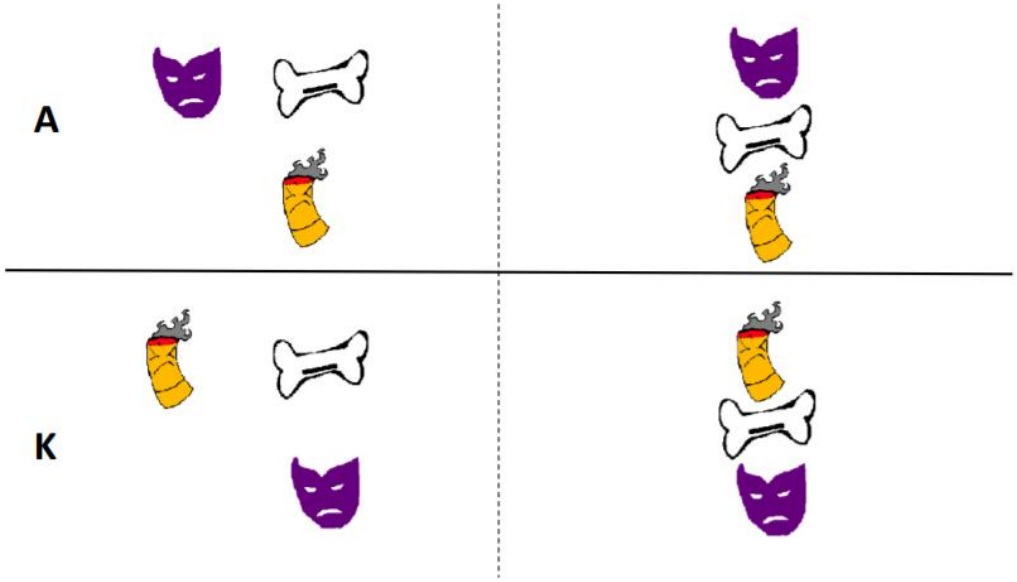


Figure 3.6: Samples of stimuli used for paired viewing. Four arrangements of the same component images are shown from lists A (top) and K (bottom). For half of the subjects list A was familiar and K was novel, whereas for the other subjects list A was novel and K was familiar.

3.2.4.3 Procedure

Table 3.1: Paired viewing training procedure

Stage	Day	Session	Image List
Pre-Training	1-2	PM	Blank (no image presented)
Training	3	PM	A-A *or* K-K
	4-5; 8-10	AM	A-A *or* K-K
		PM	AB-AB *or* KB-KB *(day 4)*; AC-AC *or* KC-KC *(day 5) etc.*
Test	11	AM	A-K *or* K-A *or* Blank

Animals were water deprived and pre-trained to go to the observing hole for juice reward for two days as follows. Disruption of the infrared beam by head entry triggered the start of a trial after a variable delay of 100-1100ms determined by a random number generator. One drop of juice was delivered after 1000ms and was followed by a period of 1000ms during which head entries were not detected. Pre-training sessions ended after the delivery of 74 drops of juice or else 20 minutes.

In the training sessions (days 3-10), trials were triggered as described above and different image arrangements were presented on each side of the screen for 4500ms, such as in Figure 3.4. A drop of juice was delivered 500ms before the end of image presentation. A subsequent trial could not be triggered for 2000ms after the end of image presentation. Thirty pairs of arrangements were presented in each session in a maximum of 20 minutes. On the first day of training, only the familiar images were

presented. After this, two sessions per day were conducted, three hours apart, for five days.

Table 3.1 describes the order of image presentation sessions. In the morning sessions, list A or K was presented each day to both eyes in random order. In the afternoon sessions, images from list A/K were presented to each eye paired with images from one of lists B to F to familiarise subjects with seeing a mixture of novel and familiar stimuli simultaneously. Half of the images in each list were presented to each side of the monitor.

In the final session, the familiar arrangements (list A or K) were presented on one screen while novel arrangements (list K or A) were presented on the other screen. The list and side of presentation for each condition were counterbalanced between subjects as shown in Table 3.2. No-image control subjects received identical training, but on test day the screen remained blank during trials. Figure 3.7 shows a schematic of the relationships between the novel, familiar, and control conditions.

Table 3.2: Counterbalancing of paired viewing conditions

Subject #	**Left Hemisphere**		**Right Hemisphere**	
	Condition	***Image Set***	***Condition***	***Image Set***
1	Familiar	A	*Novel*	K
2	Familiar	K	*Novel*	A
3	*Novel*	A	Familiar	K
4	*Novel*	K	Familiar	A
1	Control	None	Control	None
2	Control	None	Control	None

During experiments, access to water was restricted on training days with two hours ad lib access at least 30 minutes after the end of the last session. Subjects had ad lib access to water for a minimum of 24 hours in every seven days in accordance with the conditions of the project license.

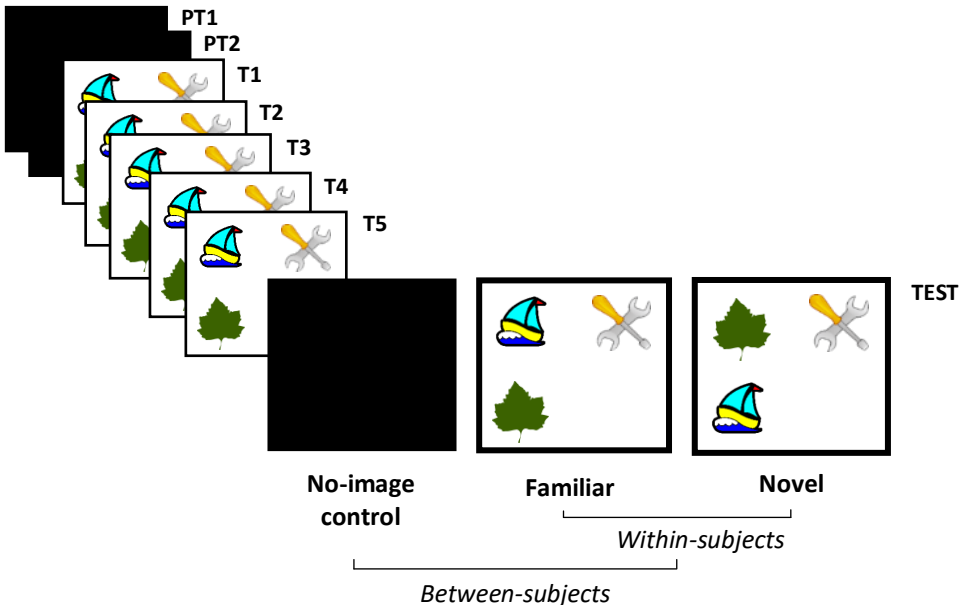


Figure 3.7: Schematic of novel, familiar and no-image control conditions of paired viewing. All subjects receive the same pre-training (PT) and training (T). In the final 'test' session, no-image control subjects are presented with a blank screen and experimental subjects are presented with the familiar set of image-arrangements on one side and the novel set of image-arrangements on the other side.

3.2.4.4 Histology

In the experiment to investigate Fos expression in the CA1 following presentation of novel and familiar image arrangements, 90 minutes after the end of the test session the subjects were transcardially perfused with 0.1M phosphate buffer (PB), followed by 4% paraformaldehyde (PFA). Brains were post-fixed in PFA for 24 hours and then transferred to 30% sucrose solution for 48 hours. Brains were frozen in liquid nitrogen and 40µm coronal sections were obtained using a cryostat.

3.2.4.5 Immunohistochemistry

The Avidin-Biotin Complex (ABC) method was used for immunohistochemistry following the protocol described in Olarté-Sanchez et al (2014). A 0.1M solution of phosphate buffered saline with 0.2% Triton X detergent (PBST) was used for all washes and to make up all solutions except where otherwise stated. All steps in the protocol were performed at room temperature except where otherwise stated.

The brain sections were washed for 10 minutes in 0.3% hydrogen peroxide solution in double distilled water (ddH_2O) to block endogenous peroxidase activity. Sections were washed four times in PBST and then incubated for one hour in 3% goat serum. The solution was replaced with a 1:10000 dilution of rabbit primary antibody to c-Fos (Synaptic Systems, 226 003) with 1% goat serum, stirred for 30 minutes and then incubated for 48 hours at 4°C. Sections were stirred for 30 minutes in the middle of the incubation period.

In the next step, the sections were washed four times and then incubated with a 1:200 dilution of biotinylated goat anti-rabbit secondary antibody (Vector Laboratories) and

1.5% goat serum for two hours. Sections were washed again four times and incubated for one hour in avidin-biotin complex (ABC) solution. Sections were washed four times in PBST and two times in 0.05M Tris-HCl buffer in ddH2O and then developed in 3,3'-diaminobenzidine (DAB) solution in ddH_2O for 1-2 minutes. The reaction was stopped using cold PB. Sections were washed in PB and later mounted on slides, then left to dry for 72+ hours. Slides were processed in ethanol and xylene and coverslips were applied using DPX mountant. Slides were visualised under a light microscope.

3.2.4.6 Cell counting

To quantify Fos expression throughout CA1, five regions of interest (ROIs) were chosen at different anterior-posterior positions relative to bregma. Two sections per ROI per hemisphere were selected from each subject. Fos-stained nuclei were counted using Cellcount4, an automated in-house cell-counting program. Parameters were adjusted by eye to maximise signal to noise ratio on sample sections and then kept constant for all images. The parameters chosen were: minimum size 15 pixels, maximum size 150 pixels, shape 3.5, and threshold 105. Raw cell counts were normalised for background staining by dividing by the mean cell count across the four images from the same subject and same ROI as described in Wan et al (1999).

3.2.5 Object in Place task

3.2.5.1 Apparatus

A square open-topped arena (95x100x50cm) with three black walls and one grey wall was used for the task. Objects made from Duplo™ were placed in each corner of the arena, 15cm from the walls. Each object was made from bricks of a different colour and had a different shape.

3.2.5.2 Procedure

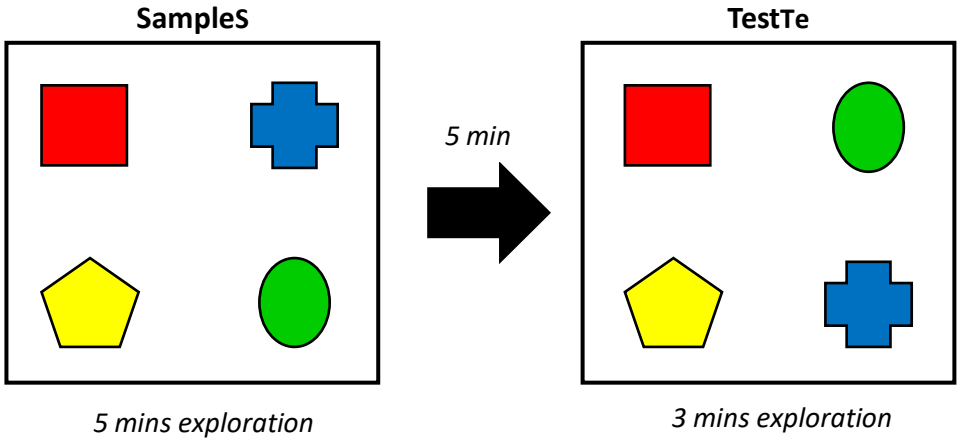


Figure 3.8: Schematic of the object- in-place test. Four objects of similar size but differing in shape and colour are arranged in an arena for subjects to explore. The test consists of a 5-minute sample phase for initial exploration, a 5-minute delay period during which the positions of the two objects on one side are switched, and a 3-minute test phase. Exploration of the objects in novel positions is compared to that of objects in familiar positions to assess object-in-place memory performance.

Prior to memory testing, animals were habituated to the empty arena for four consecutive days. On the first habituation session, cagemates were placed into the arena together for ten minutes and on subsequent sessions, animals were allowed to explore individually for five minutes.

The object-in-place task comprised a sample and a test phase separated by a delay. In the sample phase, the subjects were placed in the centre of the arena in front of the grey wall and allowed to explore the objects for 300s. Time spent exploring the objects on each side (left or right) of the arena was recorded by button presses made by the experimenter and timed by an MS-DOS script. Looking at and sniffing the object were scored as object exploration and climbing on the object whilst looking elsewhere was not. Subjects were then removed to the home cage for a five-minute delay period, during which time the objects were cleaned with ethanol to remove olfactory cues and the positions of two objects on one side of the arena were exchanged as shown in Figure 3.8. In the test phase, subjects were allowed to explore the objects for 180 seconds and the time spent exploring objects on each side of the arena was again recorded. All sessions were captured on video.

Equation 1: Object in Place task discrimination ratio

$$DR = \frac{N - F}{N + F}$$

Discrimination ratio (DR) was calculated as shown in Equation 1, where N and F represent total time spent exploring objects in novel and familiar positions respectively. If object-in-place memory is intact, subjects should spend more time exploring the objects in novel positions than those in familiar positions. The positions of the different objects and the novel side were counterbalanced between subjects.

3.2.6 Translating Ribosome Affinity Purification

Between 10 and 180 minutes after the end of the pairing viewing, animals were sacrificed by concussion and cervical dislocation, a method chosen to maximise temporal specificity and prevent confounding of gene expression data by effects of anaesthesia (see Bunting et al, 2016). Area CA1 of the hippocampus was dissected as described in section 2.2.4.2. As shown in Table 3.2, tissue was pooled from four animals to generate the experimental samples, novel and familiar, while control samples comprised tissue from both hemispheres of two animals. Tissue was maintained in homogenisation buffer during dissections of samples in the same group. (See section 2.2.4.2 for TRAP protocol.)

3.2.7 Amplification and sequencing

3.2.7.1 Comparison of amplification kits

A pilot study (by collaborator, Steven Sheardown, Takeda Cambridge) was conducted to determine the consistency of amplification of different genes in a sample of purified RNA from a cell line using two popular RNA amplification kits, the Sigma SeqPlex RNA Amplification kit (Sigma-Aldrich) and the NuGEN Ovation RNA-Seq system V2 (NuGEN Technologies Incorporated).

The consistency of amplification by the two kits was determined for nine genes with a range of abundances (Actb, App, Ctgf, Fos, Gapdh, Tfe3, Tfeb, Tgm2, and Tnks2) using qPCR. Gene expression in amplified samples prepared from 1ng and 5ng of RNA using both kits was compared with cDNA prepared from 1µg of RNA from the same sample. All qPCR samples were processed in duplicate.

Mean cycle threshold (Ct) was calculated for each gene across qPCR replicates and starting amounts for both kits and for cDNA samples. The ratios of Ct values to Actb were also calculated to compare amplification by the two kits.

3.2.7.2 Amplification and library preparation

RNA samples were amplified using Seqplex RNA amplification kit (Sigma-Aldrich) with 10ng starting material and purified using the GenElute PCR cleanup kit (Sigma-Aldrich). Amplification primers were removed and the efficiency of the primer removal reaction was confirmed by qPCR. Samples were again purified and analysed using the Agilent 2200 TapeStation (Agilent, CA).

Amplified cDNA (500ng) was prepared for sequencing by repairing the ends with T4 polymerase, extending the poly-A tails with Taq polymerase, and ligating Illumina sequencing adapters to enable sample multiplexing. DNA libraries were amplified with Phusion flash polymerase, then quantified by qPCR using NEBNext Library Quant kit (New England Biolabs). Library concentrations were calculated using average library sizes estimated from DNA ScreenTape data obtained prior to adapter ligation. Samples were pooled in groups of 5 (batches 2 and 3) or 6 (batch 1) using approximately 100ng from each sample. Pooled DNA was purified, quantified by qPCR using the NEBNext kit, and diluted to the optimum loading concentration for sequencing (10ng/µl at 2nM). Excess short fragments of DNA corresponding to the primers and adapters were removed from samples in batches 2 and 3 by gel purification at the sequencing facility following entry QC.

3.2.7.3 Sequencing

Multiplexed samples (groups of 5-6 as described above) were sequenced at GATC Biotech (Konstanz, Germany) using the Illumina Hi-Seq 2500 sequencing technology to produce 240 million 50bp single end reads per lane.

3.2.8 Bioinformatics

Reads in the first batch of sequenced samples were aligned initially using the splice-aware alignment tools Tophat2 and STAR on the Partek platform to find the approximate proportion of aligned reads in each sample and to determine the best performing tools and parameters for the analysis pipeline. Bases with a quality score of less than 20 were trimmed from the 3’ end of reads to a minimum read length of 25 prior to alignment.

3.2.8.1 Contaminant removal

A large proportion of reads were not aligned to the rat genome. The source of these reads was determined by aligning a subset of the reads not aligned to the rat genome in

sample 10.N.1.1 to the non-redundant nucleotide database (National Center for Biotechnology Information (NCBI), MD, USA) using MegaBLAST (Morgulis et al, 2008) on the Galaxy Bioinformatics platform. MegaBLAST is a version of the Basic Local Alignment Search Tool (BLAST, NCBI; Altschul et al, 1990) which performs local alignment of sequences to large databases.

3.2.8.1.1 Contaminant Assembly

Many contaminants derived from murine leukaemia virus (MLV) were found in the sequencing data and these did not align to a publicly available consensus sequence. Contaminant sequences were therefore assembled from the data using an *ad-hoc* approach summarised in Figure 3.9.

First, the reads not aligned to the rat genome from one sample (10.N.1.1) were assembled using the de-novo transcriptome assembly tool Trinity on the Galaxy platform (Galaxy beta v0.0.1 of Trinity 2.2.0, using the Bridges high-performance computing system at Pittsburgh Supercomputing Center). Next, the assembled sequences (>100) were aligned to the nucleotide database using BLASTN (NCBI) and the list of sequences was manually filtered to retain those attributed to MLV and remove rat genome sequences. A FASTA file of contaminant sequences was produced combining the assembled MLV sequences, Illumina adapter sequences, rat and mitochondrial rRNA sequences (NCBI), and the PhiX174 genome (a spike-in control added to batches 1 and 3 at the sequencing facility which was assembled by Trinity and ascribed by BLAST).

The putative MLV contaminant sequences were further refined manually with the aid of several tools. Contaminants and a published MLV reverse transcriptase sequence (AK090120.1, NCBI) were compared side-by-side using the multiple sequence alignment tool Clustal Omega (EMBL-EBI) to look for overlap and redundancy. Potential contaminants were then used as a reference to align reads from sample 10.N.1.1 and others using the short-read alignment tool Lastz (Harris, 2007) on the Galaxy platform as this tool does not require an index to be built. FastQC (v0.11.5) was used to find overrepresented sequences remaining in the unaligned reads. Overlapping overrepresented sequences were manually assembled and the source of these sequences was determined using BLAST. Sequences ascribed to MLV were incorporated into the file of contaminants, merging with overlapping MLV sequences where possible. The process was repeated until the source of all overrepresented sequences was determined and unwanted sequences were removed by inclusion in the list of contaminants.

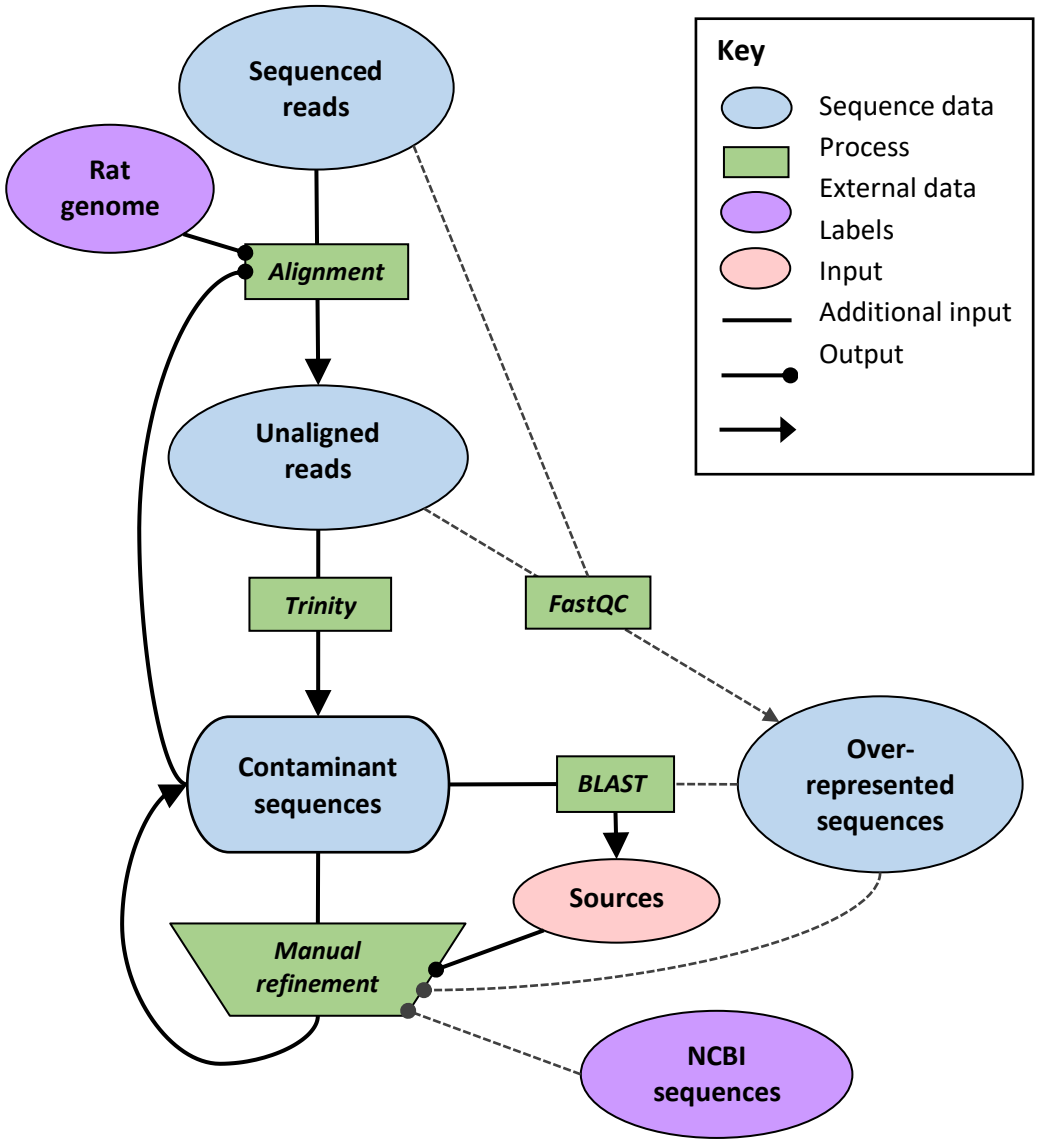


Figure 3.9 : Process for assembly of contaminant sequences from raw sequencing data. Reads not aligned to the rat genome were assembled using Trinity and assembled sequences were attributed to sources using BLAST. A FASTA file of contaminant sequences was produced from the assembled sequences and manually refined by comparison with BLAST alignments, publicly available sequences from NCBI, and overrepresented sequences in the read data. Manual refinement was aided by side-by-side comparison of sequences using the multiple sequence alignment tool Clustal Omega. Putative contaminant sequences were then used as a reference to filter reads by alignment to these sequences. If overrepresented sequences remained, overlapping sequences were combined and sources were determined using BLAST. MLV sequences were integrated into the list of contaminants. The process was repeated until no more overrepresented sequences from unwanted sources were returned by fastQC. Thick solid lines indicate key steps in the process.

3.2.8.2 Data pre-processing

The raw FASTQ files were trimmed using Trimmomatic (v0.36; Bolger et al, 2014) with the following sequence of options: first Illumina adapters were trimmed allowing a maximum of two mismatches in the seed region and an alignment score threshold of ten (recommended range 7-15 for unpaired reads); then bases with a quality score below ten were trimmed from the 3' end; next reads were trimmed to maintain an average

quality score of 20 or above across each group of four bases, and finally reads below a minimum length of 25 bases were removed.

The remaining reads were then aligned to the FASTA file of contaminant sequences using the ungapped alignment tool Bowtie2 (v2.2.9; Langmead & Sazberg, 2012). The sensitive and non-deterministic modes were used, which respectively mean that Bowtie2 prioritises sensitivity over speed and that the current time is used to initialise a pseudo-random number generator rather than the read name and sequence. The latter option was chosen due to the similarity of many of the reads. The outputs generated for each sample were two FASTQ files containing aligned and unaligned reads and a SAM file of alignments to the contaminant sequences. In the SAM files, alignments to each contaminant sequence were counted using Samtools (v1.2-41; Li et al, 2009).

3.2.8.2.1 Quality Assessment

Reads not aligned to contaminants were submitted to FastQC (v0.11.5; Andrews, 2010). The duplication rate for reads mapping to each gene was plotted against read density using the package dupRadar (Sayols et al, 2016) in R (v3.1.2).

3.2.8.2.2 Alignment and quantification

Reads not aligned to contaminants were aligned to the rat genome (Rattus_norvegicus.Rnor_6.0.dna.toplevel.fa) using the splice-aware alignment tool STAR (v2.5.1b, Dobin et al, 2013) on the BlueCrystal high-performance computing system (University of Bristol). The output statistics of Trimmomatic, Bowtie2, and STAR were collated in R and values were cross-checked between tools.

Reads were quantified to features of the rat genome using genome annotation Ensembl 89 (Rattus_norvegicus.Rnor_6.0.89.gtf) with the HTseq (v0.9.1, Anders et al, 2015) package in Python 2.7 (v2.7.6), specifying unstranded data and an alignment quality threshold of ten. The output count matrices were imported to R for further analysis.

3.2.8.3 Marker gene analysis

Lists of marker genes for hippocampal cell types and subregions were collated from the literature and CPMs of marker genes in the samples were calculated in R from the uniquely-mapped reads. Welch's t-test for samples of unequal size and with unequal variances was used to assess the significance of differences in the expression of sets of marker genes.

3.3 Results

3.3.1 Profiling associative recognition memory

3.3.1.1 Paired viewing pilot experiments

In this experiment, expression of Fos protein was analysed by counting Fos-stained nuclei in the CA1, 90 minutes following the end of the presentation of the novel and familiar arrangements. Counts were normalised for background staining as described in section 3.2.4.6. Thirty head entries were made in each session by all rats (N=8), thus 30 pairs of images were presented to each rat.

Representative photomicrographs from both novel and familiar conditions showing the CA1 region are presented in Figure 3.10A. Figure 3.10B shows the ROIs selected for analysis across the anterior-posterior extent of the hippocampus. One subject and one additional pair of sections were excluded due to poor quality, for a total of N=7 subjects and 138 images analysed. Overall, normalised expression of Fos protein was found to be significantly greater in the novel condition than in the familiar condition ($\overline{x}_{diff}$ = 0.12; SEM = 0.05) in a two-tailed paired-samples t-test at the 5% level (t = 2.22, df = 69, p = 0.0306), as shown in Figure 3.10Ci.

Normalised Fos counts at each ROI are shown in Figure 3.10Cii. Fos expression was the same in both conditions at -3.0mm from bregma and was higher in the novel condition at the ROIs with AP co-ordinates between -3.6mm and -5.2mm. However, tests for differential expression in each ROI showed that Fos counts were significantly different between conditions only in the ROI approximately 4.2mm posterior to bregma ($\overline{x}_{diff}$ = 0.25; SEM = 0.08). In this ROI, Fos counts were found to be significantly higher in the novel condition at the 5% level in a two-tailed paired-samples t-test with a Bonferroni correction for multiple comparisons (t = 3.22, df = 12, p = 0.00930).

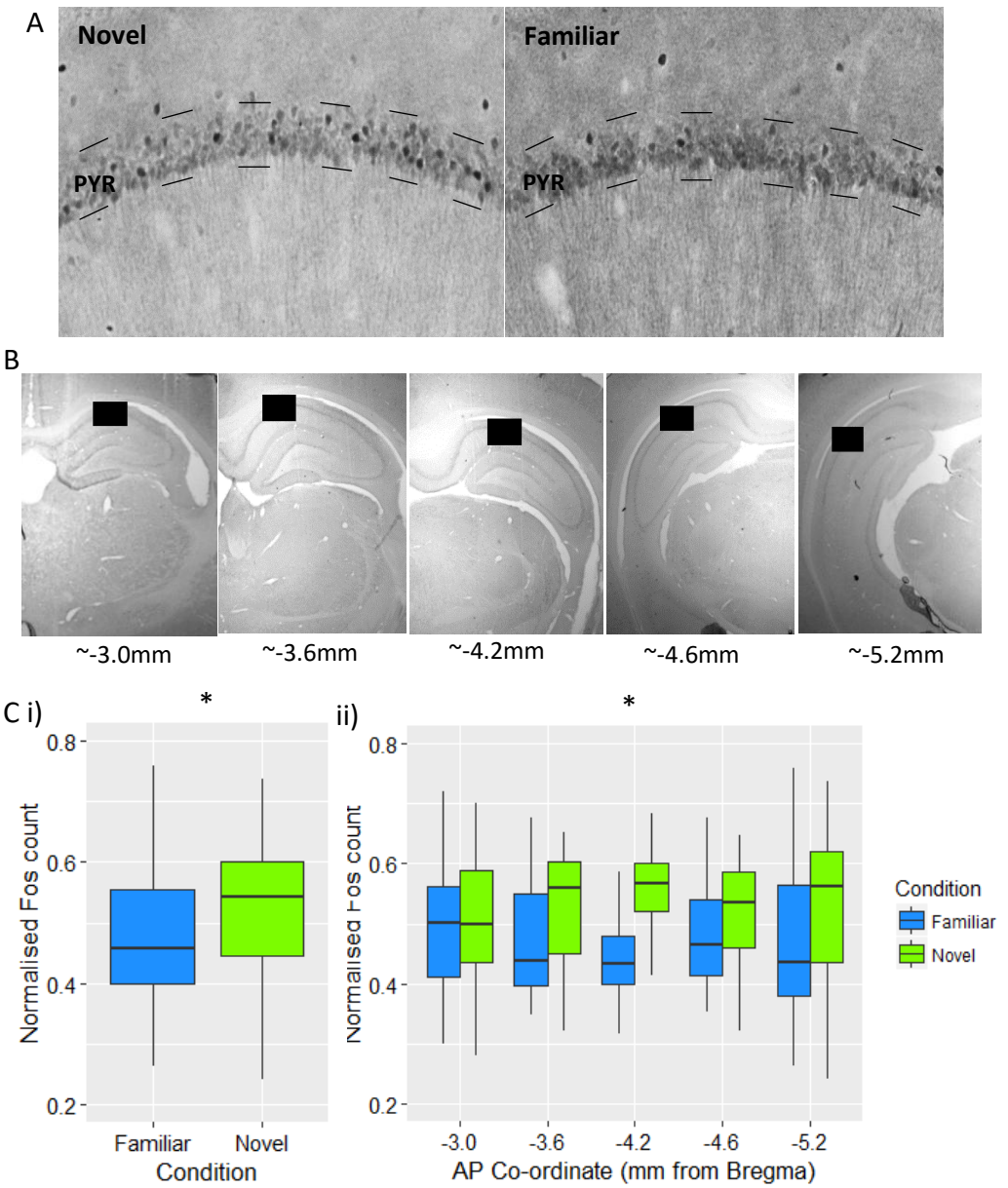


Figure 3.10: Fos expression following paired viewing of novel and familiar arrangements. A: Example of left and right hippocampal CA1 sections obtained 90 minutes after paired viewing with Fos-stained nuclei. B: Regions of interest (ROIs) for cell counting with approximate AP co-ordinates. C: Mean normalised counts of Fos-stained nuclei in CA1 sections from N=7 subjects, using 2 images for each AP co-ordinate per animal i) By condition; Fos expression was significantly higher in the novel condition at the 5% level in a two-tailed paired-samples t-test (t = 2.22, df = 69, p = 0.0306). ii) By condition and AP co-ordinate. Fos expression was significantly higher in the novel condition at the ROI -4.2mm from bregma in a two-tailed paired-samples t-test (t = 3.22, df = 12, p = 0.00930). The difference was significant at the 5% level with a Bonferroni correction for multiple comparisons.

3.3.1.2 Associative recognition memory in TRAP rats

Subjects (N=8) that had received viral injections of AAV CMV-EGFP-L10a into CA1 were tested using a single trial object-in-place task to determine whether hippocampal-dependent memory was affected by the virus. Figure 3.11 shows the mean DR and SEM (M = 0.131, SEM = 0.0872). A positive DR indicates that exploration was greater in the novel condition than in the familiar condition. However, in a one-sample one-tailed t-test, the DR was not significantly different from zero at the 5% level (t = 1.51, df = 7, p = 0.0877), indicating that more subjects are needed to confirm the trend towards discrimination between novel and familiar arrangements.

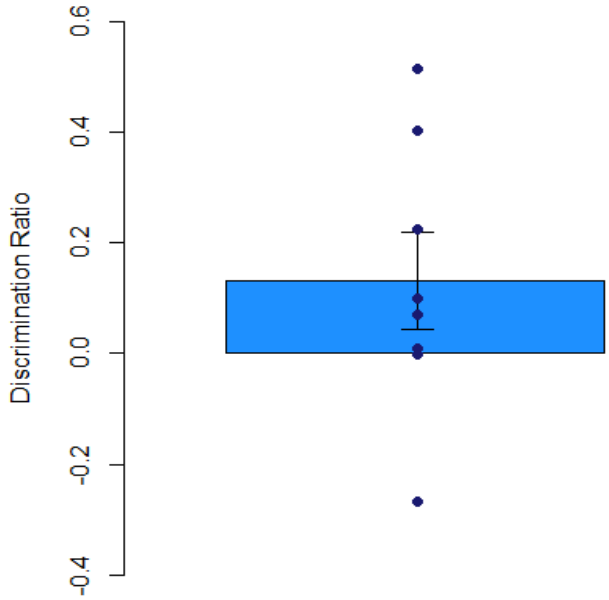


Figure 3.11: Mean discrimination ratio (DR) for exploration of items in novel vs familiar locations by rats injected with AAV CMV-EGFP-L10a into CA1. Individual subject DRs are plotted and error bar indicates standard error of the mean.

3.3.2 Preparation of RNAseq libraries

RNA samples were obtained using the TRAP method from rats trained in the paired viewing procedure at a range of timepoints following the test session.

3.3.2.1 Concentration and quality of samples

The concentrations of purified samples of immunoprecipitated RNA ranged from 0 to 100ng/µl (M = 33.46, SD = 24.61) and the mean 260/280 ratio was 2.21 (SD = 0.14). RIN^{e}s ranged from 1.0 to 7.6 (M = 5.4, SD = 2.4.). RIN^{e}s for most of the samples were in the intermediate quality range, between 5.9 and 7.6; an example trace for one of these samples is shown in Figure 3.12A. However, several samples had RIN^{e}s indicating substantial RNA degradation; an example trace for one such sample is shown in Figure 3.12B. Samples with RIN^{e}s greater than 5.0 had a minimum RNA concentration of 15 ng/µl, whereas the majority of samples with low RIN^{e}s had very low concentrations, under 10ng/µl of RNA.

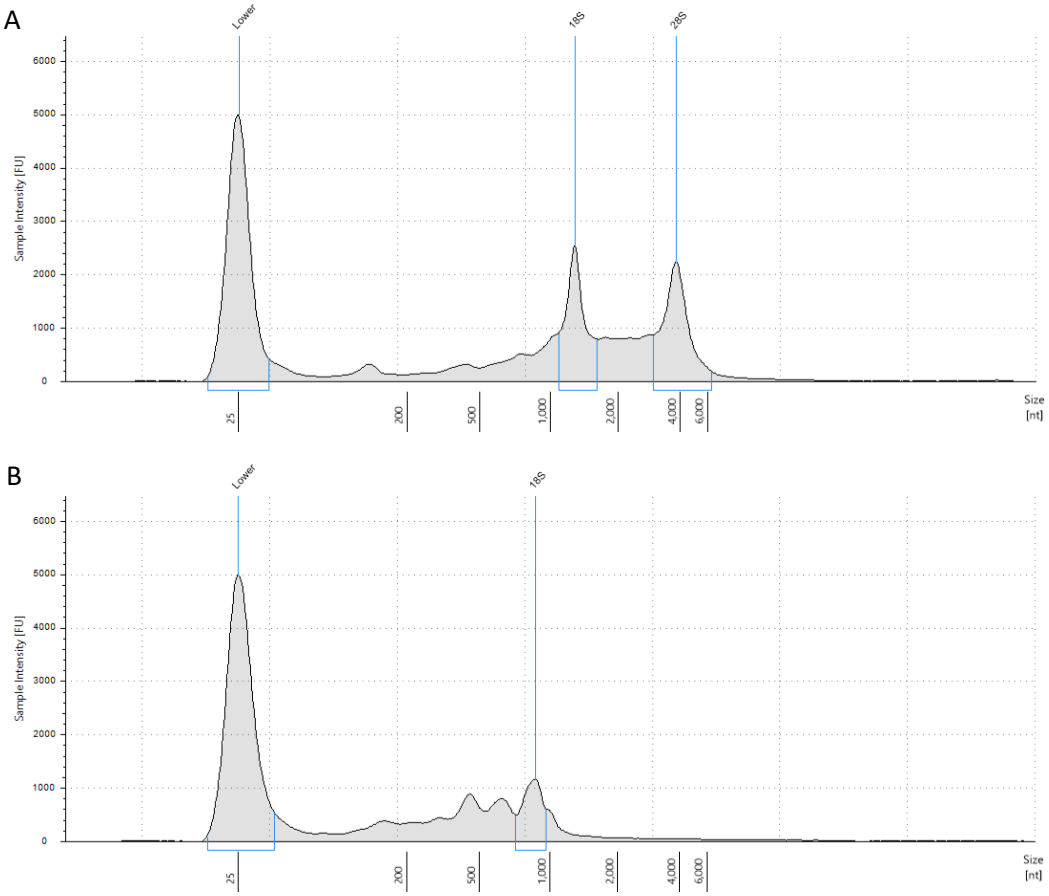


Figure 3.12: Representative electrophoretic traces from purified TRAP-generated RNA samples indicating RNA integrity.Traces were obtained using an Agilent TapeStation 2200 (Agilent, CA) and show fluorescence intensity plotted against sequence length. The peaks corresponding to 18S and 28S ribosomal RNA (rRNA) are indicated. A: An example of a sample with good quality RNA and an RNA integrity number equivalent (RIN^e) of 7.2. B: An example of a sample with poor quality RNA (RIN^e = 2.2) that was excluded from downstream analysis. In this sample, the 28S rRNA is fully degraded.

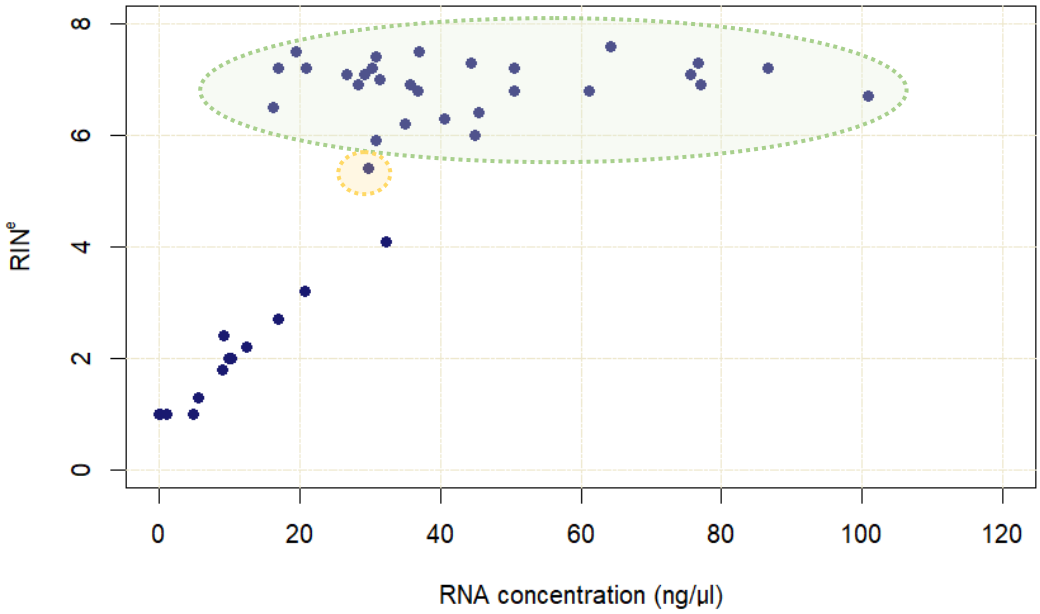


Figure 3.13: RNA concentration plotted against RNA integrity number equivalent (RIN^e) for all TRAP samples. The green ellipse marks samples retained for differential expression analysis and the yellow circle marks a sample used in some of the downstream analyses.

Figure 3.13 shows RNA concentration plotted against RIN^e for each sample. This scatter plot suggests a strong linear positive correlation between RIN^e values under 4.0 and measured RNA concentration, but no correlation between RIN^es greater than 5.0 and RNA concentrations greater than 15ng/µl. In all cases, samples prepared on the same day had similar RIN^e scores (above or below 5.0). The green ellipse marks samples retained for differential expression analysis (N = 28; RNA concentration: M = 44.43ng/µl, SD = 22.26; 260/280 ratio: M = 2.19, SD = 0.09; RIN^e: M = 6.9, SD = 0.5). Sample 30.C.10.2, circled in yellow, had a relatively low RIN^e of 5.4 and is included in marker gene analysis but not in downstream differential expression analyses presented in Chapter 4. The remaining samples (N = 13) were excluded. All excluded samples were from the first sequencing batch and these samples were not balanced across conditions. Batch effects thus must be accounted for in downstream analyses.

3.3.2.2 Comparison of commercially available RNA amplification kits

In a test to compare the consistency of amplification by the Sigma SeqPlex and NuGEN Ovation kits, the mean Ct values of genes in the samples amplified by both kits were strongly and significantly positively correlated with the mean Ct values of genes in the cDNA sample (Sigma: $r = 0.982$; NuGEN: $r = 0.932$; both $p < 0.001$, $df = 7$; 3 d.p.). However, the correlation of the Sigma samples with the control was stronger than that of the NuGEN samples.

The bar chart in Figure 3.14A shows the ratio of Ct values to Actb for genes in the cDNA control sample and in samples amplified by the two kits. Error bars indicate SD of samples prepared from different starting amounts and qPCR replicates. Figure 3.14A shows that broadly similar relationships between Ct values for the different genes were

found across the three groups, but samples amplified by the Sigma kit were more similar to the cDNA controls than samples amplified by the NUGEN kit. The ratio to Actb is lower for all genes in the NuGEN sample compared to the control except for Gapdh, which has a higher ratio. In the case of four genes (App, Ctgf, Tfe3, and Tgm2), the ratio to Actb is lower in the samples amplified by both kits than in the cDNA control and in all four cases, expression in the Sigma samples is more similar to the control than the NuGEN samples.

In Figure 3.14B, the Ct of genes in each sample is plotted against the Ct of the same genes in the cDNA control sample, taking the mean of qPCR replicates. With both kits, the Ct values and fitted linear regression models were very similar for samples prepared from different starting amounts of RNA. The mean coefficient of determination of the samples amplified with the Sigma kit is 0.963, indicating that more than 95% of the variation in the Ct values of amplified genes is explained by the abundance of these genes in the original sample. In contrast, the mean coefficient of determination of the samples amplified with the NuGEN kit is 0.865, indicating that almost 15% of the variance in Ct values was not accounted for by gene abundance in the original sample. The source of this additional variance appears to be the amplification of the least abundant genes. The Sigma SeqPlex kit was therefore used to amplify TRAP samples prior to sequencing.

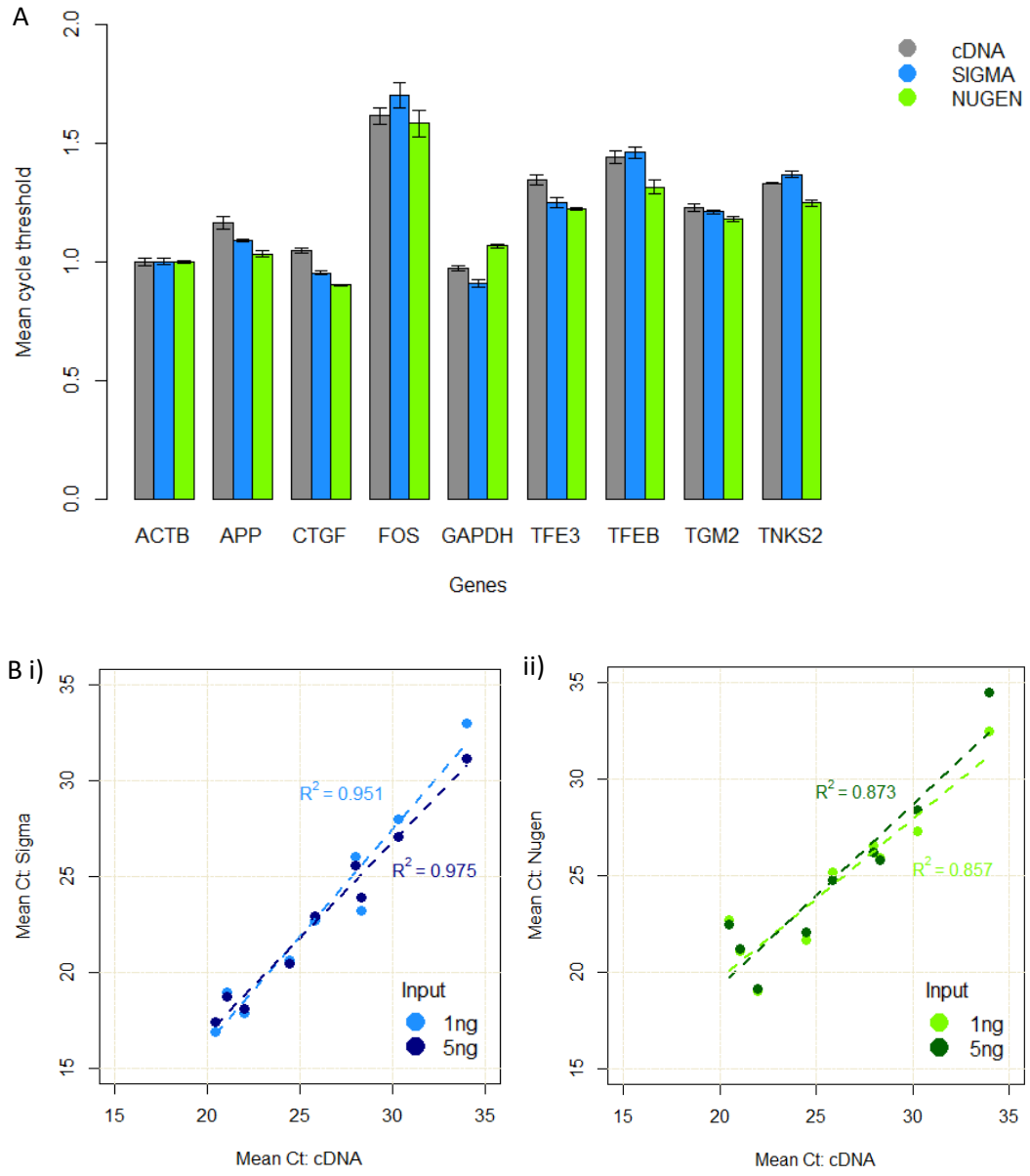


Figure 3.14: Consistency of amplification of selected genes by two commercially available amplification kits measured by qPCR. A: Ratio of mean Ct values to Actb Ct values for genes in samples amplified using each kit and for a cDNA control sample. Mean Cts are calculated from two qPCR replicates and two amplifications using different input amounts of RNA. Error bars indicate SD. B: Mean Ct of technical replicates for genes in samples amplified from 1ng and 5ng of starting material compared to cDNA controls. A linear regression model is fitted to each series and the coefficient of determination for each regression is indicated. i) Sigma SeqPlex amplification kit. ii) NuGEN Ovation RNA-Seq system V2.

3.3.2.3 Amplified DNA

Following amplification and reverse transcription, DNA samples were analysed using the Agilent 2200 TapeStation (Agilent, CA). The mean length of DNA fragments across the three batches prior to adapter ligation was 240.52 (SD = 12.13).

3.3.3 Quality control of sequencing data

3.3.3.1 Quality assessment of sequenced samples

Quality assessment of the raw sequencing data was conducted using fastQC. All values are reported to the degree of precision given by fastQC and representative examples of the fastQC output are presented in Figure 3.15 and Figure 3.16 from sample 180.N.3.1, a sample of mid-range quality (excluding samples with low RINs).

The sequencing quality was on average very high at all positions within reads (mean and median quality scores greater than 30). The mean quality score of reads in all samples in the first two batches sequenced was 35, corresponding to a probability of an incorrect base call of less than 0.01%. In the third batch, the mean quality was 39 for all samples in lane one and 36 for samples in lane two. All samples had negligible N content and all sequences had length 51. The number of reads per sample was highly variable, ranging from 17.2M to 49.2M (M = 33.75M, SD = 7.57M).

The mean GC content of the samples with $RIN^e > 5.0$ was 51% (SD = 1), whereas the mean GC content of the low RIN samples was 54%. Excluding the low RIN samples, read count was moderately and significantly negatively correlated with GC content ($r = -0.642$, $p < 0.001$, $df = 27$). There were biases in sequence content across all positions within reads and the overall GC content of reads did not fit the expected normal distribution, as shown in the examples in Figure 3.15A and Figure 3.15B respectively.

Samples also had a high proportion of duplicated sequences. In sample 180.N.3.1 for example, 81.66% of reads were duplicates and some reads were duplicated more than 10,000 times (Figure 3.16A). However, there is a smooth and gradual decline in the proportion of reads in sample 180.N.3.1 with between 1 and 9 copies, suggesting that there is a diversity of genes with low expression counts and amplification in this region is not extreme. Just over 50% of distinct sequences in this library were not duplicated and reads duplicated more than 50 times account for a very small proportion of distinct sequences. Figure 3.16B shows that there are positional biases towards specific short sequences. Overrepresented sequences were reported in all samples, as could be inferred from the sequence duplication plots, but many more overrepresented sequences were found in samples with RIN^e below 3.0 (>100).

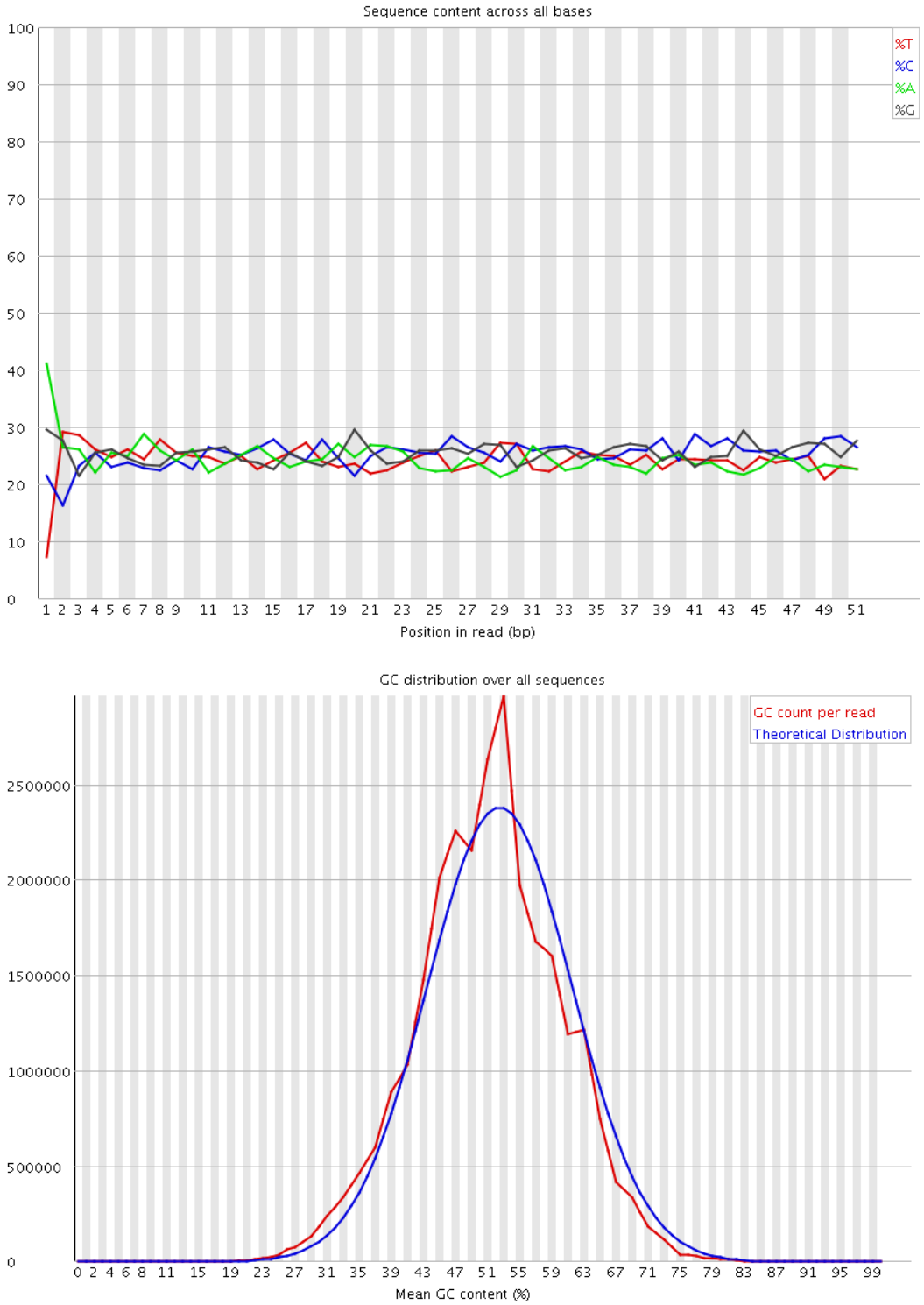


Figure 3.15: Representative examples of FastQC graphs of sequence content. A: Proportion of reads containing each of the four bases at each position. Base content is weakly biased at all positions. B: Distribution of mean GC content across all reads (red) compared to a normal distribution with the same mean (blue). The observed distribution does not fit the theoretical distribution.

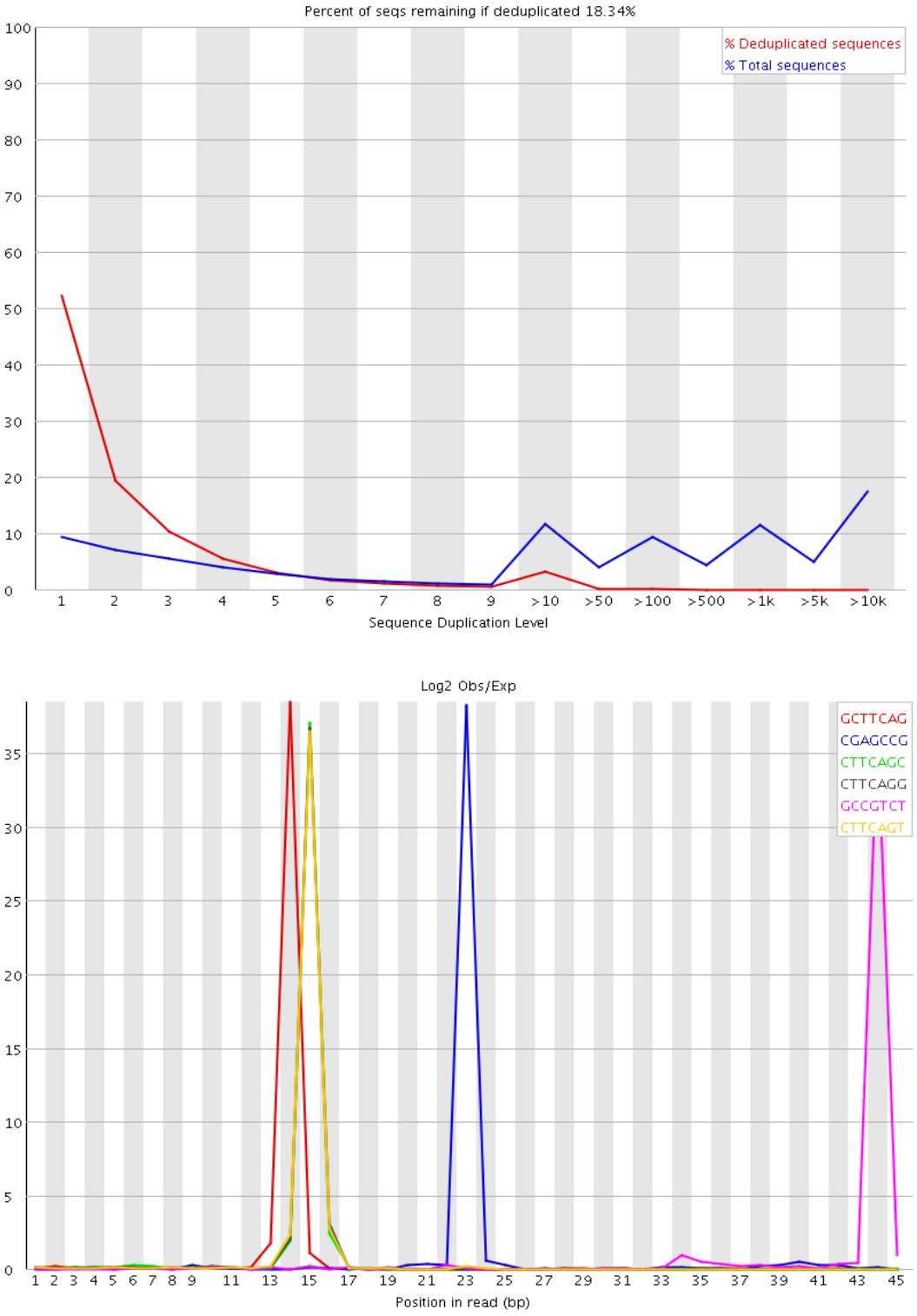


Figure 3.16: Representative examples of FastQC graphs of sequence duplication and k-mer content. A: The percentage of all sequences (blue) and the percentage of distinct sequences (red) are plotted at each duplication level. Deduplicated sequences account for 18.34% of reads. Sequences duplicated more than ten times are divided into bins of uneven size. B: The $\log_2$ ratio of observed k-mer count to expected count according to the base content of the library for the top six most overrepresented k-mers. Biases towards different k-mers are observed at different positions.

3.3.3.2 Optimisation of alignment

The first batch of sequenced samples were aligned using the splice-aware alignment tools Tophat2 and STAR on the Partek platform to determine the proportion of aligned reads in each sample and optimise the analysis pipeline. Prior to alignment, 4-6% of reads per sample were trimmed and under 2% of reads were discarded.

Alignment to the rat genome using TopHat2 with default parameters showed two distinct patterns of alignment: samples with a very low (<5%) proportion of unique alignments and a large proportion of multiple alignments (~75%), and samples with approximately 30% unique alignments, 30% multiple alignments and 40% unaligned. Figure 3.17 shows that when alignment statistics are plotted against RIN^e, but not RNA concentration, two distinct clusters corresponding to these alignment patterns are formed. Samples with low RIN^es (<5.0) have a very low proportion of unique alignments but a high proportion of total alignments. A minimum RIN^e of 5.0 was therefore set as a threshold for the analysis and subsequent experiments.

The output of Tophat2 was compared to that of the splice-aware alignment tool STAR. More reads in the samples retained for downstream analysis were found to be aligned and uniquely aligned by STAR and fewer reads were aligned by STAR in the samples with low RIN values (see Figure 3.18 for examples). STAR was therefore used for subsequent alignments.

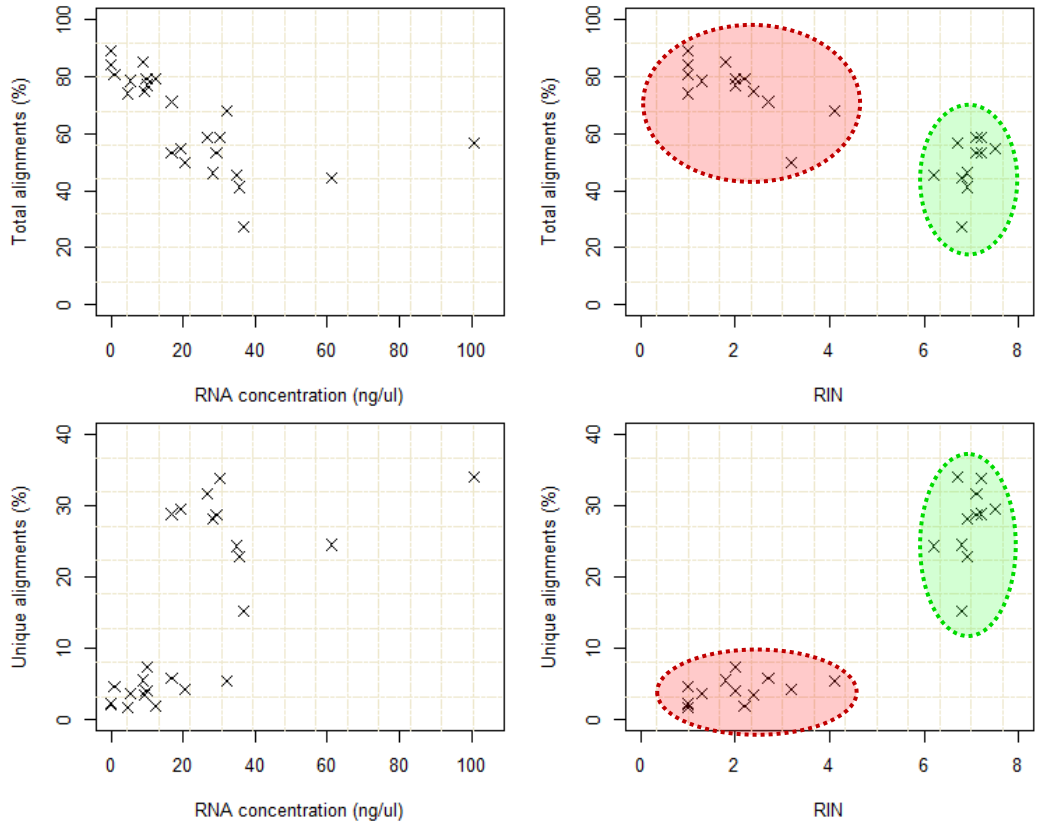


Figure 3.17: Effects of RNA integrity number (RIN) and RNA concentration on alignment. Scatter plots showing the relationships between input RNA statistics (RNA concentration and RIN) and the proportion of aligned reads using Tophat2 (total and unique alignments). Samples with RIN^e<4.5 (circled in red) were found to have qualitatively different alignment statistics from samples with RIN^es between 6.0 and 8.0 (circled in green). Samples with low RIN^es had a low proportion of unique alignments and high proportion of total alignments.

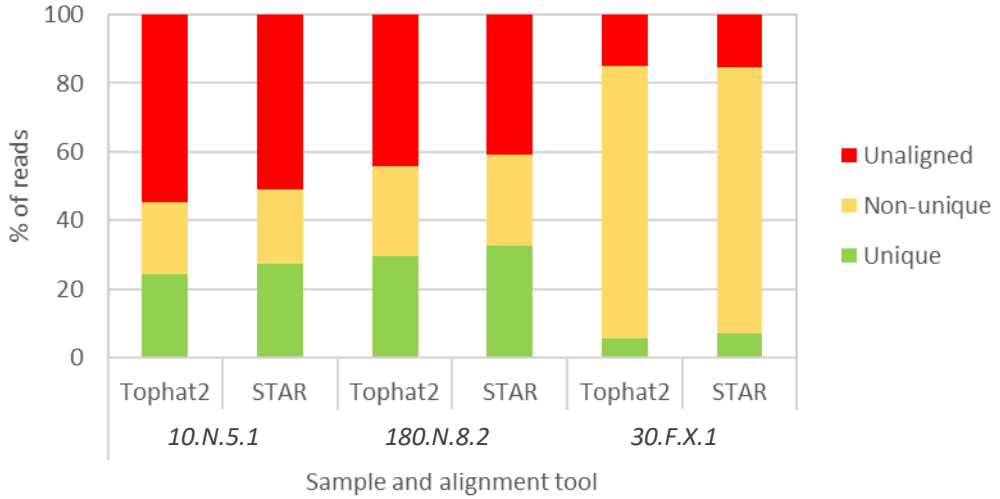


Figure 3.18: Comparison of Tophat2 and STAR splice-aware alignment tools. Stacked barchart showing the percentages of uniquely aligned, multiply aligned, and unaligned reads by Tophat2 and STAR for samples from batches 1 (10.N.5.1) and 2 (180.N.8.2) and a low RIN sample from batch 1 (30.F.X.1).

Two unbound samples were sequenced in batch 3 (90.F.15.3 and 90.N.15.3) for comparison with IP samples. Sequencing data was pre-processed and aligned to the rat genome using STAR as described above. Figure 3.19 shows that more than 95% of reads were aligned in the unbound samples and ~65% of reads were uniquely aligned, whereas only 60-80% of reads were aligned in the corresponding IP samples and less than 50% were uniquely aligned.

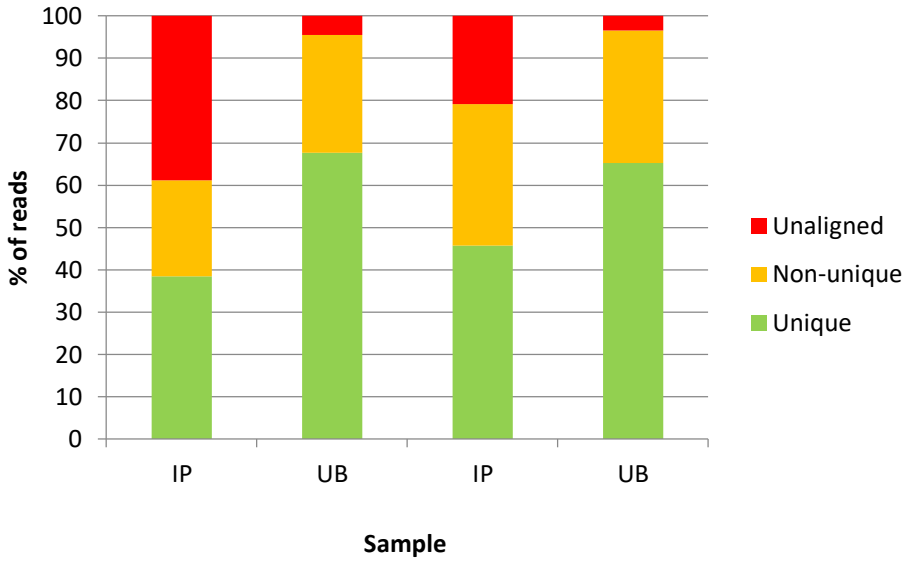


Figure 3.19: Percentage of aligned reads in unbound RNA samples compared to immunoprecipitate samples. A greater proportion of reads were aligned and uniquely aligned in unbound (UB) samples compared to the corresponding immunoprecipitate (IP) samples.

3.3.3.2.1 Origins of contaminant sequences

MegaBLAST was used to determine the source of overrepresented sequences and reads not aligned to the rat genome. From a subset of unaligned reads in sample 10.N.1.1 (N ≈ 2.1 million), 38 million alignments to 71,724 GenInfo identifiers (GIs) were found, accounting for just over 80% of the sample of unaligned reads. Over 98% of hits (i.e. alignments) were accounted for by just 0.69% of unique GIs and fourteen GIs had over one million hits each, accounting for almost 50% of hits. These fourteen GIs were associated with murine leukemia virus (MLV), indicating contamination. Contaminant sequences were assembled as described in section 3.2.8.1.1.

MLV sequences were not found in significant numbers in the unbound RNA samples, despite that the IP and unbound samples were processed in parallel and contained similar starting amounts of RNA.

3.3.3.2.2 Quality assessment following contaminant removal

Quality assessment was repeated following read trimming and the removal of contaminant sequences. Quality control statistics are presented in Table 3.3 and representative examples of fastQC graphs are shown in Figure 3.20 and Figure 3.21 for sample 180.N.3.1. This sample was chosen because the proportion of sequences remaining after deduplication (the most variable of the fastQC statistics after pre-processing) was close to the dataset mean.

In all samples, the sequence GC content distribution fit the theoretical distribution following contaminant removal (example in Figure 3.20A), although mean GC content was reduced compared to the original samples (M=48, SD=1). The per base sequence content was close to uniform from base 20 onwards, as in Figure 3.20B. However, nucleotide biases remained in the first 20 bases of reads and the k-mer “CTTCAG” was strongly overrepresented at position 15 in all samples, as in Figure 3.20B and Figure 3.21B. “CTTCAG” matches the universal sequence at the 5’ end of the SeqPlex amplification kit primers. Another k-mer, CGCTAAA, was overrepresented at position 20 in some samples.

Table 3.3: Quality control statistics for sequenced RNA samples used for analysis

Sample ID	*Input RNA*		Total # Reads (millions)	*QC post-trimming and contaminant removal*				
	Conc. (ng/ul)	RIN[e]		% GC	% Dedup seqs	# Over-rep. seqs	*dupRadar* Intercept	*dupRadar* Slope
10.N.1.1	19.4	7.5	33.3	50	40.0	0	0.30	2.42
10.F.1.1	16.8	7.2	34.1	50	38.4	0	0.33	2.35
180.C.2.1	30.3	7.2	31.7	49	38.3	0	0.23	2.54
180.N.3.1	26.7	7.1	27.0	49	35.0	0	0.42	2.22
180.F.3.1	29.2	7.1	17.2	49	36.9	0	0.45	2.25
10.C.4.1	100.9	6.7	23.8	48	36.4	0	0.43	2.32
10.N.5.1	35.0	6.2	23.3	49	40.9	0	0.29	2.37
10.F.5.1	61.1	6.8	21.1	51	42.7	0	0.32	2.39
30.C.6.1	28.2	6.9	25.3	49	46.9	0	0.27	2.51
30.N.7.1	36.7	6.8	29.8	50	41.1	1	0.39	2.23
30.F.7.1	35.6	6.9	27.8	49	35.3	0	0.48	2.19
180.N.8.2	50.4	6.8	35.7	49	47.2	0	0.12	2.93
180.F.8.2	86.7	7.2	35.7	48	34.4	3	0.26	2.53
10.N.9.2	44.9	6.0	38.6	47	40.8	1	0.26	2.53
10.F.9.2	30.9	5.9	45.2	46	25.4	1	0.73	2.21
30.C.10.2	29.8	5.4	28.6	47	28.1	1	0.59	2.18
30.N.11.2	45.5	6.4	43.4	48	41.0	1	0.18	2.83
30.F.11.2	40.6	6.3	49.2	47	38.3	0	0.25	2.61
90.C.12.2	75.7	7.1	47.9	46	39.8	0	0.19	2.81
90.N.13.2	20.9	7.2	45.5	48	35.1	2	0.21	2.76
90.F.13.2	76.8	7.3	47.4	48	34.0	3	0.25	2.60
90.N.14.3	77.0	6.9	30.3	49	27.9	0	0.32	2.54
90.F.14.3	31.4	7.0	33.6	48	29.9	1	0.33	2.56
90.N.15.3	30.8	7.4	33.0	49	28.0	0	0.23	2.68
90.F.15.3	37.1	7.5	39.2	48	28.9	1	0.24	2.68
10.N.16.3	16.3	6.5	39.8	49	26.2	1	0.45	2.40
10.F.16.3	50.5	7.2	39.8	49	31.8	0	0.35	2.50
90.N.17.3	64.3	7.6	43.5	48	27.2	0	0.38	2.49
90.F.17.3	44.4	7.3	34.7	48	30.8	0	0.29	2.57

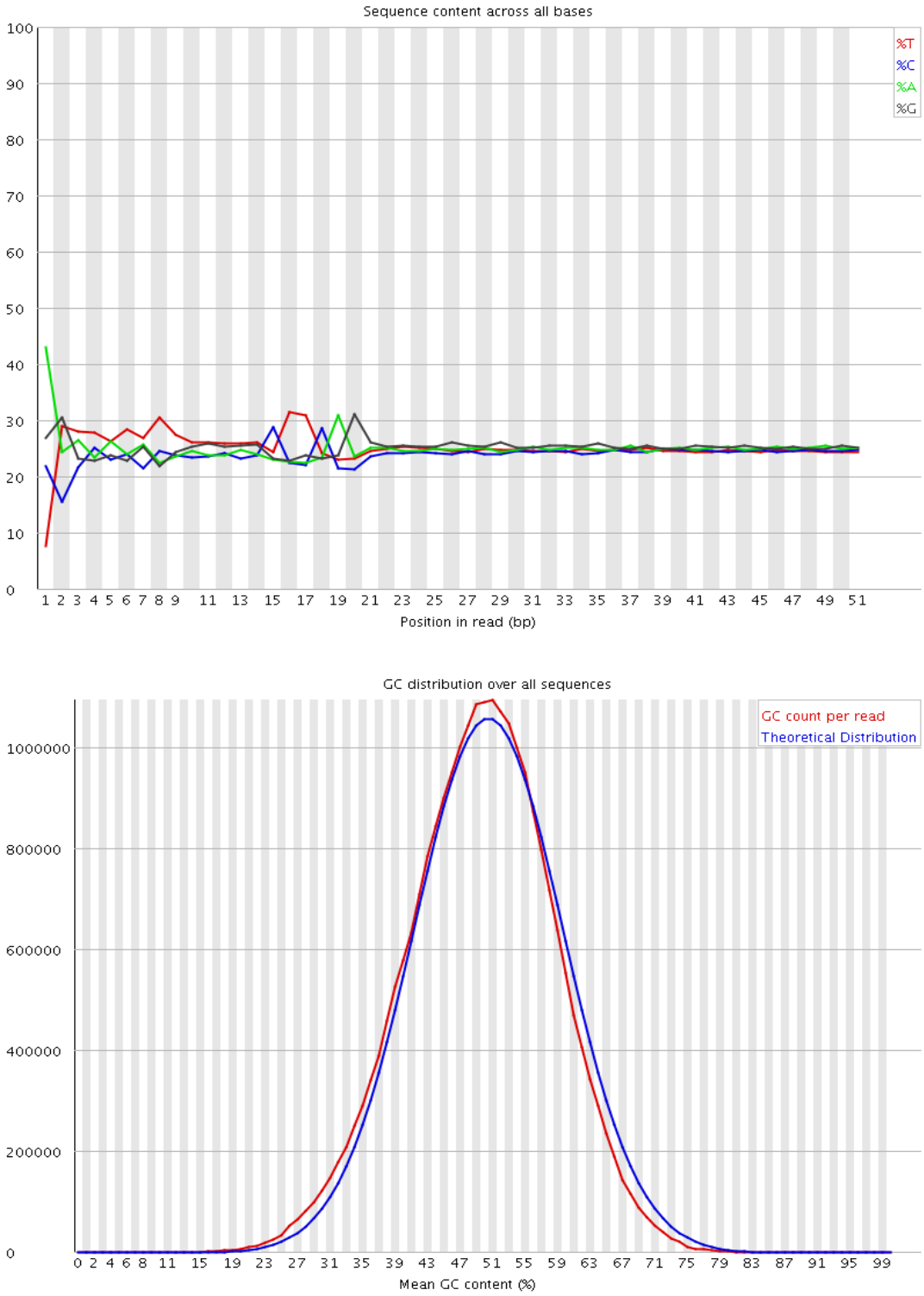


Figure 3.20: Representative examples of FastQC graphs of sequence content after pre-processing. A: Proportion of reads containing each of the four bases at each position. Biases are evident across the first 20 positions. B: Distribution of mean GC content across all reads (red) compared to a normal distribution with the same mean (blue). The observed distribution fits the theoretical distribution.

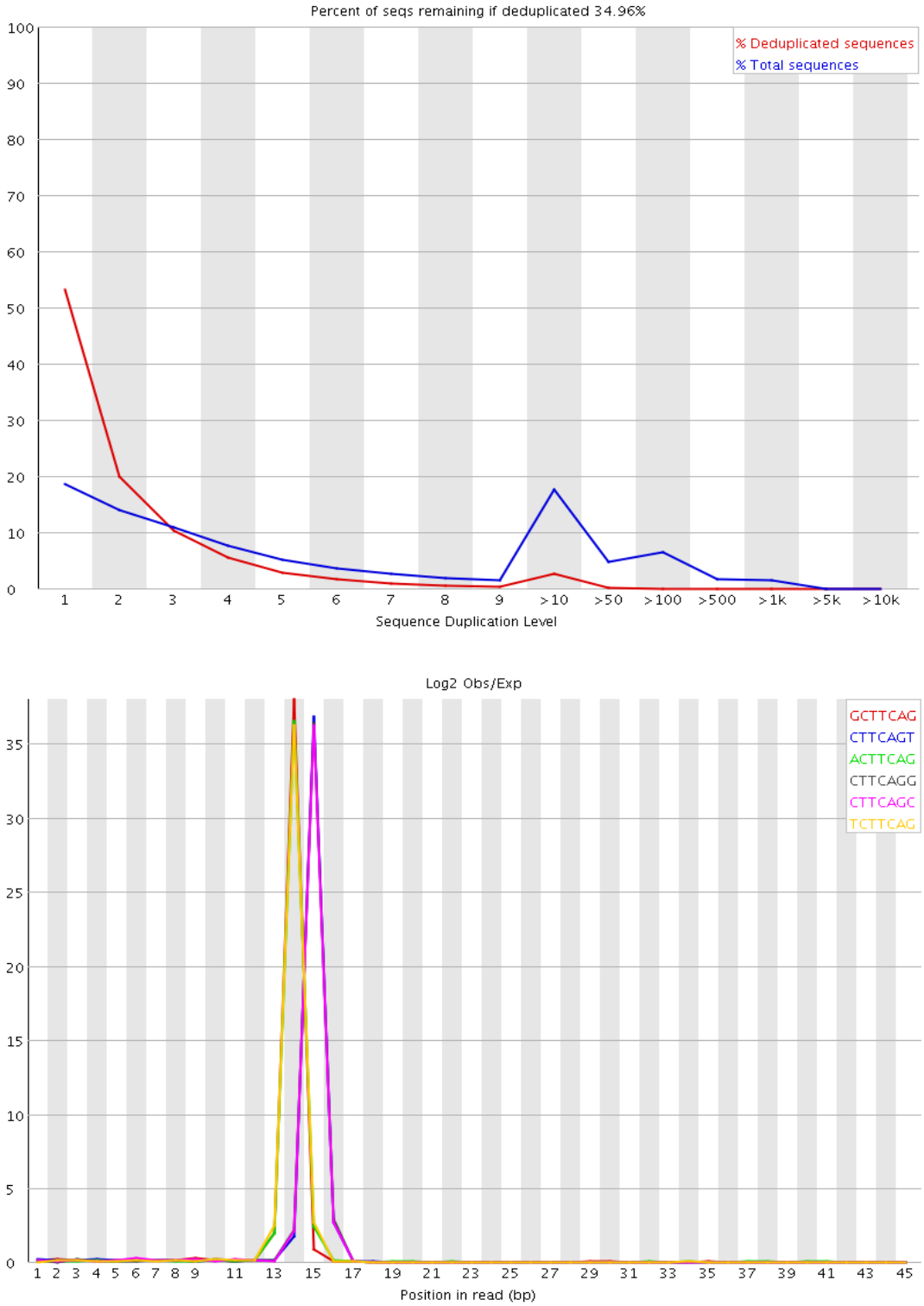


Figure 3.21: Representative examples of FastQC graphs of sequence duplication and k-mer content after pre-processing. A: The percentage of all sequences (blue) and the percentage of distinct sequences (red) are plotted at each duplication level. Deduplicated sequences account for 34.96% of reads. Very few sequences are duplicated more than 500 times. B: The $\log_2$ ratio of observed k-mer count to expected count according to the base content of the library for the top six most overrepresented k-mers. Biases are observed towards k-mers containing CTTCAG at position 15.

Six overrepresented sequences were found in eleven samples, shown in Table 3.4. Two of the overrepresented sequences, each reported in only one sample, aligned to highly abundant rat genes, ubiquitin B (UBB) and calmodulin 2 (Calm2). The remaining four

sequences were found to originate from MLVs but had mismatches compared to the previously assembled contaminants. These MLV sequences were each specific to a batch of samples prepared at the same time.

Table 3.4: Overrepresented sequences and sources

Samples	Sequence	Source
30.N.7.1	GACCCCACCATCAGGCTTAGCCAGCTAACTGCAGTAA CGCCATCTTGCAAG	MLV
180.F.8.2, 10.N.9.2, 10.F.9.2, 30.C.10.2, 90.N.13.2, 90.F.13.2	CTAAGAACTTAGAACCTCGCTAAAAAGGACCCTACAC CGTCCTGCTGACCA	MLV
180.F.8.2, 30.N.11.2, 90.N.13.2, 90.F.13.2	AGGACGGTGTAGGGTCCTTTCCAGCGAGGTTCCAAGT TCTTAGTCTGGTGC	MLV
180.F.8.2	CTTTATCCTGGATCTTGGCCTTCACGTTCTCGATGGTG TCACTGGGCTCCA	Rat UBB
90.F.13.2	CAAAGAAGCTTTCTCCCTATTTGACAAGGACGGGGAT GGGACAATAACAAC	Rat Calm2
90.F.14.3, 90.F.15.3, 10.N.16.3	ATTGTGCTTATAGACCCCAGAGTTCAGGTCAGGTAGA AAGAATGAATAGAA	MLV

After pre-processing, heavily duplicated reads (>500 copies; see Figure 3.16) were removed and the proportion of unique sequences remaining was hence increased (M = 35.40%, SD = 6.11), as shown in Table 3.3. Figure 3.21 shows that uniquely occurring sequences in sample 180.N.3.1 accounted for more than 50% of distinct sequences and just under 20% of all sequences after pre-processing.

The duplication rate for reads mapping to each gene was plotted against read density using dupRadar (Sayols et al, 2016). Logistic regression models fitted to the samples had a mean intercept of 0.33 (S.D. = 0.13) and mean slope of 2.49 (S.D. = 0.20), representing the baseline level of duplication and the progression of duplication rate with read density respectively (see Table 3.3). Intercept and slope were strongly and significantly negatively correlated ($r = -0.842$, $p < 0.001$, $df = 27$), meaning that the proportion of duplicated reads rose more steeply with read density in samples with lower baseline amplification. However, duplication rate was correlated with read density in all samples. Figure 3.22 shows examples of plots indicative of varying degrees of amplification. Intercept was moderately but significantly negatively correlated with RIN ($r = -0.448$, $p = 0.015$, $df = 27$) and the intercepts of the two samples with $RIN^e < 6.0$ were outliers, i.e. more than two standard deviations greater than the mean, indicating that baseline amplification was higher in these samples.

RNAseq data analysis pipelines commonly include a deduplication step to avoid errors due to nonlinear PCR amplification (Hashimoto et al, 2014; Klepikova et al, 2017). However, a high proportion of duplicate reads in RNAseq data occur naturally, particularly in single-end read datasets (Bansal, 2017), and recent papers have concluded that computational methods of deduplication do not improve and may infact worsen accuracy and false discovery rate, in addition to greatly increasing the false

negative rate (Hashimoto et al, 2014; Parekh et al, 2016; Klepikova et al, 2017). Thus a deduplication step was not included in the pipeline.

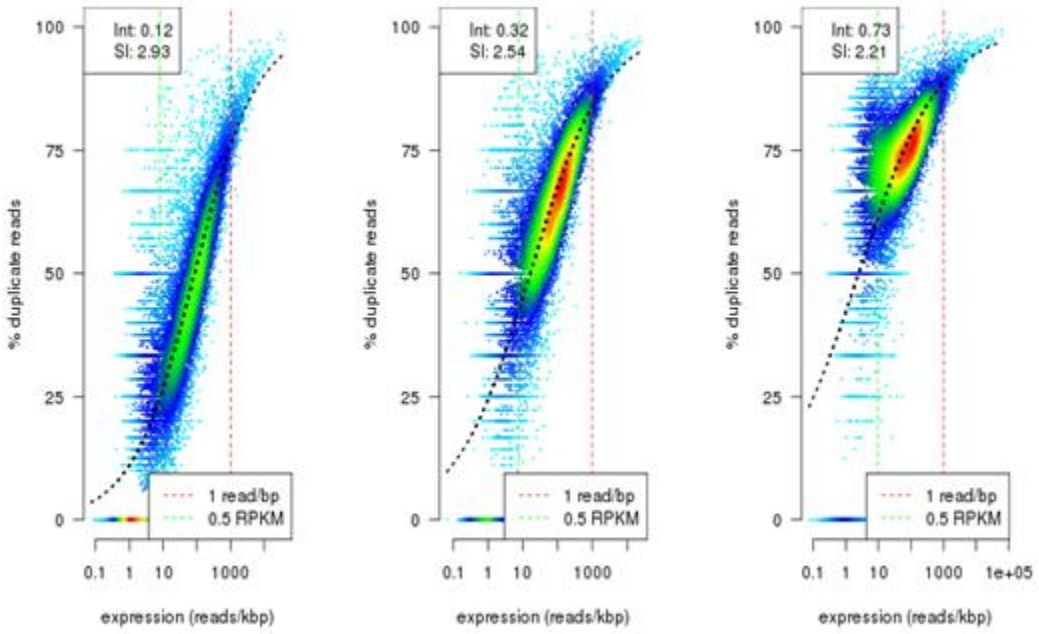


Figure 3.22: Representative dupRadar plots of duplication rate against read density of genes in a sample. From left to right, examples of plots with intercept and slope values suggestive of relatively low (180.N.8.2), close to the mean (90.N.14.3), and high (10.F.9.2) levels of amplification among the sequenced samples. Intercept and slope values are in the top left corner of each plot.

3.3.4 Library composition

Mapping statistics were analysed to determine the proportion of reads in each sample that were trimmed, aligned to known contaminants, aligned to the rat genome, or unaligned, as shown in Figure 3.23.

Figure 3.23A shows that there was substantial variation in the total number of reads per sample (see also Table 3.3), but similar proportions of reads were aligned to each source across samples. Effects of sequencing batch on sample composition were also observed. Samples in batch 1, for example, contained a higher proportion of trimmed reads compared to batches 2 and 3, and samples from batch 2 had more total reads, on average. One sample (30.N.7.1) had a very high proportion of trimmed reads compared to the other samples, suggestive of adapter contamination. Up to 40,000 additional reads aligned to adapter sequences were removed from samples in the Bowtie2 contaminant removal step (M = 6433), but these reads accounted for a very small proportion of each sample.

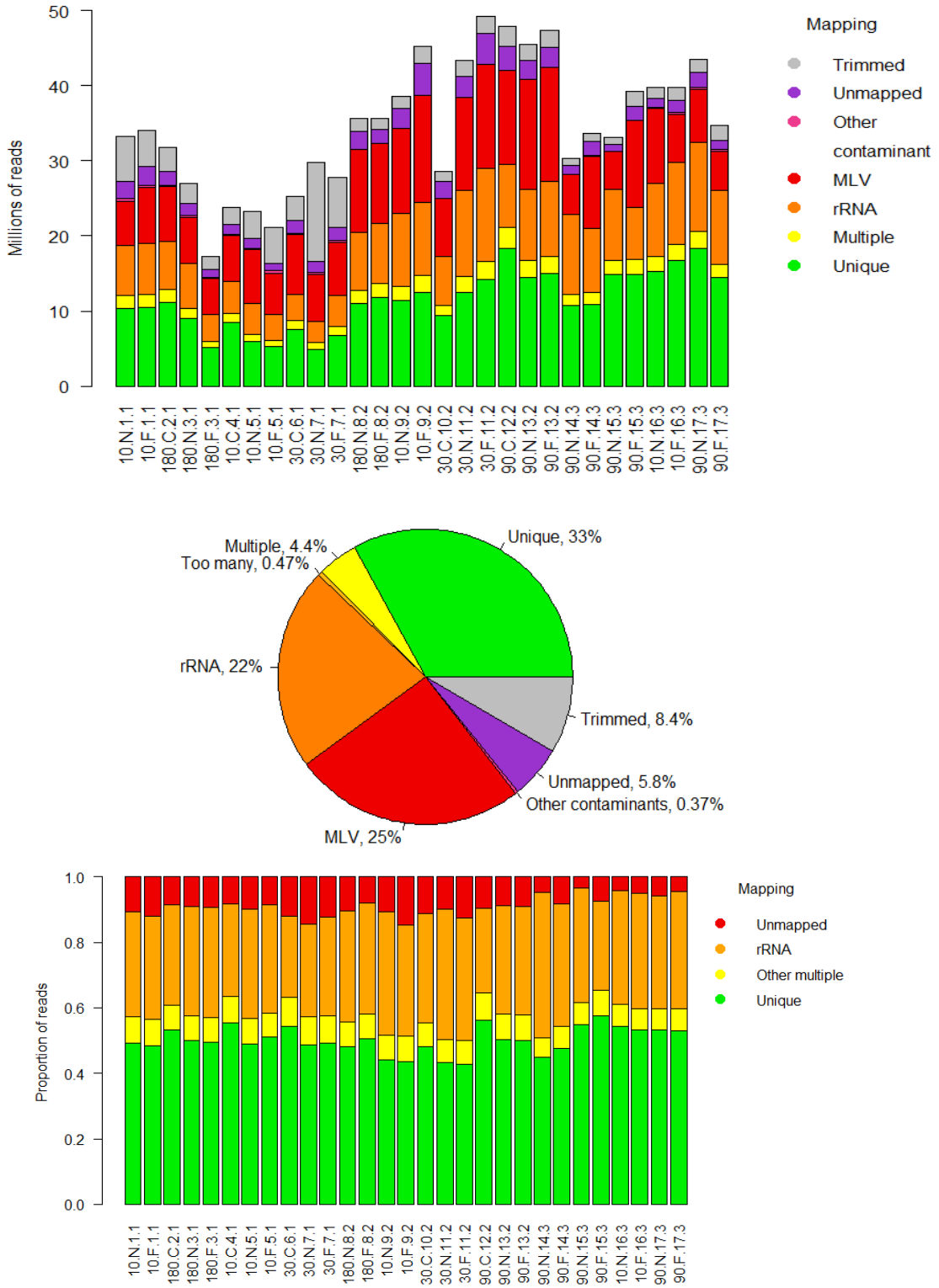


Figure 3.23: Alignment of reads to the rat genome and other sources. A: Total number of reads in millions in each sample and the number of reads trimmed or mapped to different sources. B: Piechart displaying the percentage of reads in all samples mapped to different sources. C: Proportion of trimmed reads in each sample mapped to the rat genome by Bowtie2 (rRNA) and STAR (unique and multiply aligned) after filtering for adapter sequences, the PhiX174 spike-in control, and MLV-derived contaminants.

The percentage of all reads trimmed, aligned to each source, or unaligned is shown in Figure 3.23B. Sixty percent of reads were aligned to the rat genome (including rRNA) and one third of all reads were uniquely aligned. MLV contamination accounted for 25% of reads and rRNA accounted for just over 20%. A small percentage of reads were unmapped (5.8%).

Figure 3.23C shows that after trimming reads and filtering out contaminants from other sources (adapters, MLV, and PhiX), the proportion of mapped reads was similar in all samples. Samples from batch 3 had a slightly lower proportion of unaligned reads compared to samples from the first two batches. Overall, after pre-processing, more than 90% of reads were aligned to the rat genome and ~50% were uniquely aligned.

3.3.5 Expression of hippocampal cell type markers

3.3.5.1 Neurons and glia

The expression of cell type marker genes (listed in Table 3.5) was analysed to provide an indirect indication of the distribution of transduced cell types in the TRAP RNA samples. Figure 3.24A shows boxplots of CPMs for selected marker genes. The neuronal marker, Map2, is by far the most strongly expressed and expression of Map2 mRNA was more than tenfold higher than expression of most of the interneuron and glial markers. CaMKiiα is the second most strongly expressed marker gene, followed by the other markers of excitatory neurons, Slc17a7 (vGlut1), Grin2a (NMDAR2A) and Grin2b (NMDAR2B). Expression of the excitatory neuron marker Slc17a6 (vGlut2) is negligible and this gene has been found previously to have very low expression in CA1 pyramidal neurons (Herzog et al, 2006).

Among the markers of other neural cell types, the most strongly expressed is the interneuron marker gene Gad2. Genes found in parvalbumin (Pvalb) interneurons (PVINs) and astrocytes were expressed in the samples, whereas expression of markers of oligodendrocytes and the interneuron subtype marker 5HT3A was negligible. The expression level of a third interneuron subtype marker somatostatin (Sst) was between that of parvalbumin and 5HT3A. Expression of S100b is particularly strong compared to the other astrocytic markers, which may be explained by expression of this gene in some CA1 pyramidal neurons (Rickmann & Wolff, 1995). The expression of key cell type markers was reasonably consistent across samples (see Supplemental Figure 3.26).

As shown in Figure 3.24B, the mean $\log_2$ CPM values of pyramidal neuron markers expressed in CA1 (CaMKiiα, Slc17a7, Grin2a, Grin2b) were found to be significantly greater than those of PVINs (Gad1, Gad2, Pvalb; $t = 3.05$, $df = 4.17$, $p = 0.036$) and astrocytes (Gfap, Slc17a2, Slc17a3, S100b; $t = 4.00$, $df = 5.77$, $p = 0.008$) in Welch two-sample t-tests. The mean expression of PVIN and astrocyte markers did not differ significantly.

Table 3.5: Cell type markers

Gene	Full name	Cell type(s)	Source
Map2	Microtubule-associated protein 2	Neurons	Abcam, 2016
CaMKiiα	Calcium/calmodulin-dependent protein kinase II alpha	Excitatory neurons	Ainsley et al, 2014
Slc17a7	Solute carrier family 17 member 7		
Slc17a6	Solute carrier family 17 member 6		Abcam, 2016
Grin2a	Glutamate ionotropic receptor NMDA type subunit 2A		
Grin2b	Glutamate ionotropic receptor NMDA type subunit 2B		
Gad1	Glutamate decarboxylase 1	Inhibitory neurons	
Gad2	Glutamate decarboxylase 2		
Pvalb	Parvalbumin	Parvalbumin INs	Rudy et al, 2011
Sst	Somatostatin	Somatostatin INs	
Htr3a	5-hydroxytryptamine receptor 3A	5HT3A INs	
Gfap	Glial fibrillary acidic protein	Glia	Abcam, 2016
Slc1a3	Solute carrier family 1 member 3	Astrocytes	
Slc1a2	Solute carrier family 1 member 2		
S100b	S100 calcium binding protein B		
Olig1	Oligodendrocyte transcription factor 1	Oligodendrocytes	
Olig2	Oligodendrocyte lineage transcription factor 2		
Sox10	SRY box 10		

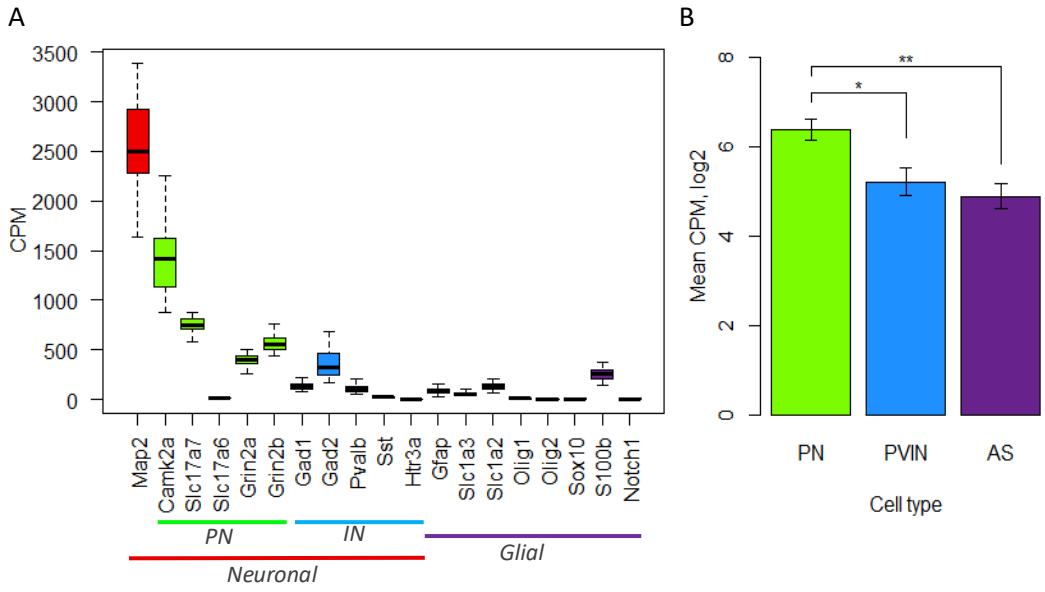


Figure 3.24: Expression of neural cell type markers within TRAP samples. A: Boxplots indicating median, IQR and range of marker gene CPMs. Markers found in neurons, excitatory and inhibitory neurons, and glia are indicated below the x axis and by colour (see Table 3.5). B: Mean log2 CPM of genes expressed in CA1 pyramidal neurons (PN), parvalbumin interneurons (PVINs), and astrocytes (ASs). Error bars indicate SE of mean CPMs of genes in each cell type. Expression of PN markers was significantly different from markers of PVINs (t = 3.05, df = 4.17, p = 0.036) and ASs (t = 4.00, df = 5.77, p = 0.008) in Welch two-sample t-tests.

3.3.5.2 Hippocampal subregions and CA1 domains

The expression of genes reported in the literature to be markers of hippocampal subregions and CA1 domains (see Table 3.6) was analysed to confirm the specificity of dissections and viral transduction.

Figure 3.25A shows that the CA1 subregion-specific marker gene Wfs1 (Takeda et al, 2001), which has been reported to be specific to the dorsal domain in mice (dCA1; Dong et al, 2009), was by far the most strongly expressed gene. All genes with mean CPMs greater than 50 were reported markers of CA1 or dCA1. The expression of markers of other hippocampal subregions was low and expression of putative markers of the ventral domain of CA1 (vCA1) was negligible. The expression of selected subregion and domain marker genes was consistent across samples (see Supplemental Figure 3.27).

Gpt2, reported by Dong et al (2009) to be expressed in both intermediate CA1 (iCA1) and dCA1, also had very low expression, whereas the other reported iCA1 and dCA1 marker Zbtb20 was expressed more strongly, at a similar level to most of the other dCA1 markers. Markers of CA1 domains are not necessarily specific to the CA1 subregion within the hippocampus.

Figure 3.25Bi shows that the expression of CA1 marker genes was significantly higher than expression of markers of other hippocampal subregions in a Welch two-sample t-test ($t = 4.49$, $df = 4.31$, $p = 0.009$). Overall, CA3 and DG markers were expressed at similarly low levels (CA3: $M = 15.4$, $SEM = 4.6$; DG: $M = 15.6$, $SEM = 4.8$), whereas markers expressed in CA2 had higher expression on average ($M = 44.49$, $SEM = 5.4$). However, all but one of the CA2 markers overlapped with the lists of CA3 and DG markers, as indicated in Table 3.6 and Figure 3.25A. Figure 3.25Bii shows that expression of dCA1 markers was significantly higher than expression of vCA1 markers in a Welch two-sample t-test ($t = 4.68$, $df = 8.98$, $p = 0.001$).

Table 3.6: Markers of hippocampal subregions and CA1 domains

Gene	Full name	Subregion(s)	Source
Bhlhe40	Basic helix-loop-helix family, member e40 (Bhlhb2)	CA1	Lein et al, 04
Etv1	Ets variant 1		Lein et al, 04[†]
Wfs1	Wolframin ER transmembrane glycoprotein	CA1; CA1d	Takeda et al, 01*; Dong et al, 09; Kohara et al, 14
Nov	Nephroblastoma overexpressed		Zhao et al, 01[†]; Lein et al, 04; Dong et al, 09
Kcnh7	Potassium voltage-gated channel subfamily H member 7	CA1d	Dong et al, 09
Ndst4	N-deacetylase and N-sulfotransferase 4		
Zbtb20	Zinc finger and BTB domain containing 20	CA1d/CA1i	Dong et al, 09
Gpt2	Glutamic--pyruvic transaminase 2		
Dcn	Decorin	CA1v	Dong et al, 09
Grp	Gastrin releasing peptide		
Htr2c	5-hydroxytryptamine receptor 2C		
Col5a1	Collagen type V alpha 1 chain		
Gpc3	Glypican 3		
Rgs14	Regulator of G-protein signaling 14	CA2	Kohara et al, 14
Spock1	Sparc/osteonectin, cwcv and kazal like domains proteoglycan 1 (Testican)	CA2/CA3	Zhao et al, 01; Lein et al, 04
Elavl2	ELAV like RNA binding protein 2		Lein et al, 04
Socs2	Suppressor of cytokine signaling 2		
Bok	BOK, BCL2 family apoptosis regulator	CA3	Zhao et al, 01[†]; Lein et al, 04
Prkcd	Protein kinase C, delta		

Frzb	Frizzled-related protein	CA2/CA3/DG	Lein et al, 04
Ntf3	Neurotrophin 3 (NT-3)	CA2/DG	Zhao et al, 01[†]; Lein et al, 04
Tiam1	T-cell lymphoma invasion and metastasis 1		
Pcp4	Purkinje cell protein 4		
Dsp	Desmoplakin	DG	Lein et al, 04
Ppp1r1a	Protein phosphatase 1, regulatory inhibitor subunit 1A		
Mcm6	Minichromosome maintenance complex component 6		
Igfbp5	Insulin-like growth factor binding protein 5		Zhao et al, 01[†]; Lein et al, 04
Trpc6	Transient receptor potential cation channel, subfamily C, member 6 (Trp6)		Zhao et al, 01; Lein et al, 04
Nrep	Neuronal regeneration related protein (P311)		Lein et al, 04

*Restricted expression reported in specified hippocampal subregion(s) or CA1 domain(s) in mouse brain unless otherwise indicated; *Rat brain; [†]Enriched expression.*

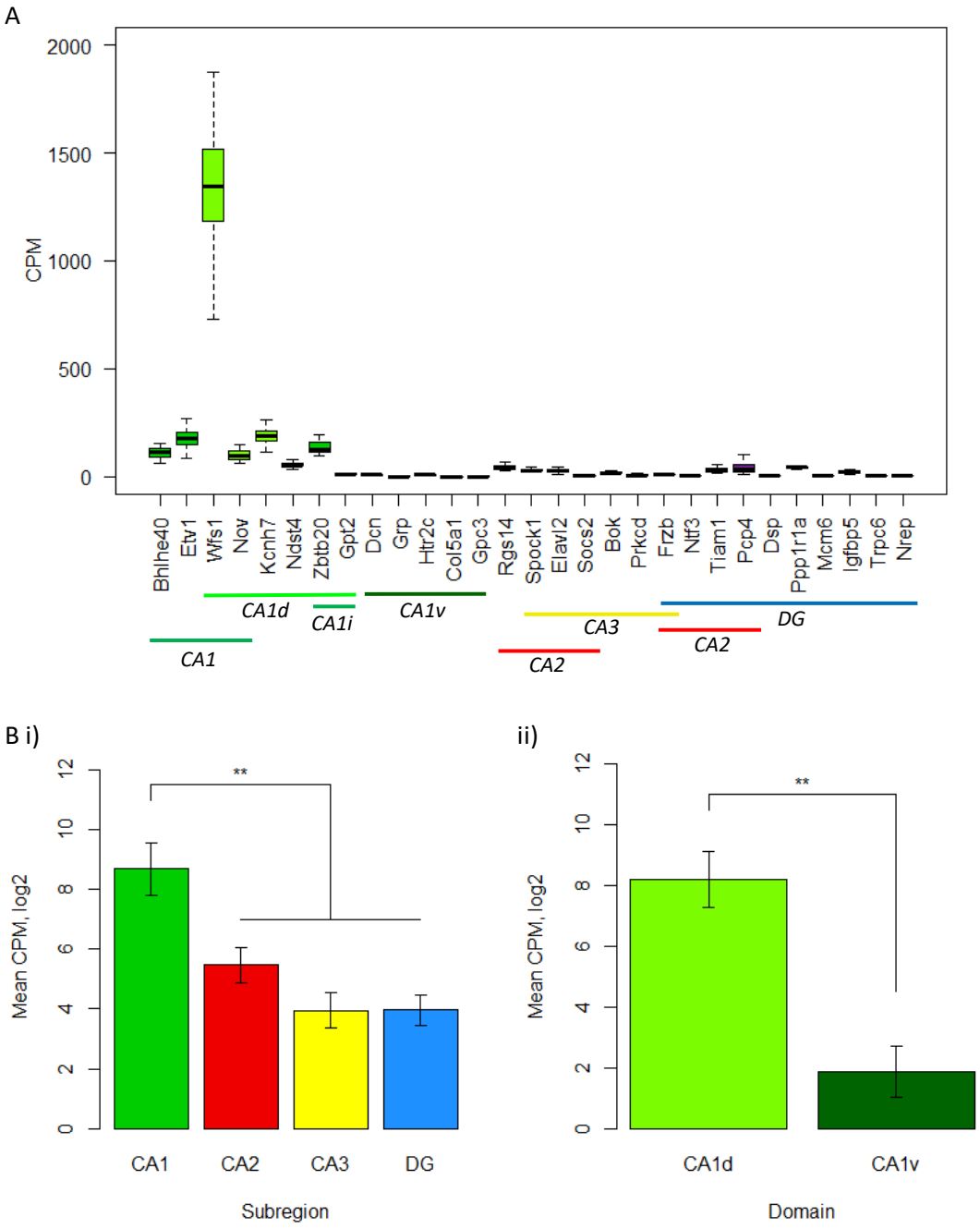


Figure 3.25: Expression of markers of hippocampal subregions and CA1 domains in sequenced samples. A: Boxplots of counts per million (CPMs) of individual marker genes across all samples. Hippocampal subregions and CA1 domains are indicated below the x axis and by colour (as in Table 3.6). B: Mean log_2 CPMs of genes enriched in different areas of the hippocampus (as shown in A and listed in Table 3.6). Error bars indicate standard error of mean CPMs of genes in each group. i) Hippocampal subregion markers. Overlapping genes are included in CA2, CA3 and dentate gyrus (DG) groups. Expression of CA1 markers was significantly different from markers of other hippocampal subregions in a Welch two-sample t-test (t = 4.49, df = 4.31, p = 0.009). ii) CA1 domain markers. Expression of dorsal (CA1d) markers was significantly greater than expression of ventral (CA1v) markers in a Welch two-sample t-test (t = 4.68, df = 8.98, p = 0.001). CA1i, intermediate CA1.

3.4 Discussion

In this chapter, the methods for an experiment to profile the translatome of CA1 neurons during associative recognition memory formation are presented, with particular focus on the preparation and quality control of RNAseq libraries. The methods and their suitability for the experiment are evaluated below.

3.4.1 Experiments to profile associative recognition memory

3.4.1.1 Paired viewing

In the paired viewing pilot experiment, significantly higher Fos expression was found in the CA1 in the novel compared to the familiar condition, thus replicating the result of Wan et al (1999). The effect size was smaller than that found by Wan et al, which may be due to a small sample size and imperfect counterbalancing due to the loss of one subject from the final analysis. The difference in expression was found to be significant in dorsal hippocampus approximately 4.2mm posterior to bregma, within the region transduced by the AAV CMV-EGFP-L10a virus using the injection protocol optimised in Chapter 2.

Previous studies of Fos expression in rat hippocampus following behavioural protocols involving changes in item locations have reported differential expression at varying anterior-posterior positions in CA1. In the paired viewing study by Wan et al (1999), the AP co-ordinate was -5.2mm from bregma and in another study, differential expression of Fos was found in dCA1 at -3.24mm following an object-location task (Mendez et al, 2015), but these studies did not report expression data at other locations within CA1. In the present study, expression was higher in the novel condition at these positions, but the difference was not significant, suggesting that the lack of an effect may be due to low experimental power. Low experimental power relative to effect size may also explain inconsistencies in the literature: following cue rearrangement in the radial arm maze, Jenkins et al (2004) observed differential expression of Fos in CA1 at -2.85mm from bregma but not at -5.0mm, whereas following exposure to novel objects in the bowtie maze, Fos was found be differentially expressed in vCA1 at -4.8mm from bregma, but not in iCA1 at the same AP co-ordinate, nor in dCA1 at -2.5mm (Albasser et al, 2010b), yet Fos expression was higher in the novel than the familiar condition in every one of these cases.

3.4.1.2 Object-in-place task

The virus, which tagged a ribosomal protein with EGFP, was not expected to affect memory performance (Warburton et al, 2005). However, as the paired viewing procedure provides no direct method of assessing associative recognition memory performance, a subgroup of animals injected with the TRAP virus into CA1 were tested in an object-in-place task. Although subjects explored objects in novel positions more than those in familiar positions, the discrimination ratio was not significantly greater than zero. As only a single object-in-place trial was conducted, an effect may have been obscured by the unfamiliarity of the task and lack of training. Additional subjects and/or training sessions are needed to confirm intact memory performance.

3.4.2 Quality control of TRAP RNA samples

3.4.2.1 Input RNA

The concentration of purified immunoprecipitated RNA was variable, which may be caused by multiple factors including the efficiency of homogenisation, immunoprecipitation, or RNA purification. Some of the samples had very low RINs, indicating RNA degradation, which may affect the retention of differential expression patterns (Shanker et al, 2015). However, low RINs in this experiment were found in conjunction with very low concentrations in most cases, well below the range for which RIN values can be reliably calculated (Mueller et al, 2016). In addition, RIN values are calibrated for total RNA, meaning that the relationship between electrophoretic profiles and RNA integrity may be different for TRAP RNA samples. This possibility was suggested by Kratz et al (2014), who found that RINs differed significantly between samples of ribosome-bound RNA from fractions corresponding to different ribosomal populations; for example, they reported a mean RIN of 3.3 for RNA associated with endoplasmic reticulum-bound ribosomes. In the present study, quality and alignment statistics were found to be comparable across samples with RIN^{e}s between 6.0 and 8.0, with a minimum RNA concentration of 15ng/µl. These data were used for differential expression analysis. In samples with RIN^{e}s below 5.0 in contrast, very few reads were uniquely aligned to the rat genome. In support of this threshold, Schroeder et al (2006) found when validating the RIN algorithm that samples with a range of RINs formed two distinct clusters, as opposed to having a graded relationship, when the expression of housekeeping genes measured by qPCR was plotted against RIN.

3.4.2.2 Optimisation of amplification protocol

Genes with different expression levels were found to be amplified consistently using the SeqPlex (Sigma-Aldrich) amplification kit. The Ovation V2 (NuGEN) kit has been used in recent TRAP studies (Ainsley et al, 2014, Nectow et al, 2017) and similar results were found for the two kits in a recent comparison study (Shanker et al, 2015). In this experiment however, expression measurements in samples amplified using the Ovation kit were less strongly correlated with the original samples, particularly for low abundance genes, suggesting that amplification biases were present. The SeqPlex kit was therefore used to amplify samples in the present experiment. This amplification method amplifies rRNA less effectively than mRNA and analysis of library composition showed that rRNA accounted for just over 20% of total reads.

3.4.2.3 Detection of murine retroviral sequences

The NGS data was found to be contaminated by sequences from murine leukaemia virus. These contaminants may have been introduced during library preparation as the reverse transcriptase enzyme is derived from MLV (Rittié & Perbal, 2008) and MLV-like and other murine retroviral sequences have been reported as contaminants in PCR reagents (Sato et al, 2010; Tuke et al, 2011; Zheng et al, 2011), nucleic acid extraction columns (Erlwein et al, 2011) and human cell lines (Zheng et al, 2011). The high proportion of reads attributed to MLV contaminants may be due to an impure reverse transcriptase preparation containing multiple copies of MLV transcripts that were then amplified.

However, more than 95% of reads in sequenced samples of unbound RNA from this experiment aligned to the rat genome and a previous TRAP study similarly found that substantially fewer reads aligned to the mouse genome in IP samples than in input RNA samples (Ainsley et al, 2014). These results suggest that contaminants were introduced at the immunoprecipitation step, the most likely source of which is the antibodies to EGFP as these were derived from mouse. Furthermore, a mouse monoclonal antibody used to modulate the activity of Taq polymerase has previously been reported as a source of MLV contamination (Sato et al, 2010), and high expression of endogenous retroviruses has been reported in rodent cell lines used for antibody production (Shepherd et al, 2003). The sequences of MLV-like contaminants were assembled from the data and reads mapping to these sequences as well as known sequencing contaminants were filtered from the data prior to alignment to the rat genome. This step was performed to prevent potential biases due to sequence similarity between contaminants and the rat genome.

3.4.3 Quality of sequencing data

The quality of sequencing was very high according to all measures, although the number of reads per sample was highly variable. Nucleotide biases were identified at the beginning of reads, as commonly occurs in RNAseq data due to biases in sampling the pseudo-random sequence primers during reverse transcription (Hansen et al, 2010). Analysis of sequence duplication showed that unique reads were underrepresented in the samples and reads mapped to genes with low read density were duplicated, indicative of technical duplication from the amplification step (Sayols et al, 2016). The baseline duplication rate was moderately correlated with RNA integrity and highest in the two samples with RIN^es below 6.0. However, duplication rate was proportional to read density in all samples, enabling detection of biological variation in gene expression. The observations of variation in library size and batch effects illustrate the importance of correction for these factors in downstream differential expression analyses.

3.4.4 Expression of hippocampal marker genes

Analysis of the expression of marker genes for neural cell types, hippocampal subregions, and CA1 domains in the sequenced samples of TRAP RNA showed that the most abundant of these transcripts were markers of the target cell population, CA1 pyramidal neurons. Several marker genes were chosen for each cell type of interest as the expression levels of different genes are highly variable.

Of the neural cell type markers, the neuronal marker Map2 was by far the most strongly expressed in the TRAP samples and the markers of excitatory neurons were more strongly expressed than markers of interneurons and glia, indicating that pyramidal neurons were the most abundantly transduced cell type. These results suggest that interneurons and astrocytes were transduced alongside excitatory neurons, in accordance with the observation of EGFP-expressing cells outside the pyramidal cell layer in section 2.3.3.2. As interneurons account for only ~10% of CA1 neurons (Bezaire & Soltesz, 2013), it is likely that these cells were strongly transduced. Of the markers of interneuron subtypes, parvalbumin was most strongly expressed. Astrocytes on the

other hand are highly abundant (Nedergaard et al, 2003). The expression of S100b was stronger than of the other astrocytic markers, which may be explained by expression of this gene in some hippocampal neurons (Rickmann & Wolff, 1995). Interestingly, S100b expression has been found to be upregulated in CA1 during LTP (Sokolova et al, 2009) and a role of this gene in learning may also have contributed to its relatively high expression. Expression of markers of oligodendrocytes was very low and the presence of small quantities of mRNA from cell types that were not transduced may be explained by non-specific binding of mRNA to the magnetic beads during immunoprecipitation (Doyle et al, 2008).

Among the markers of hippocampal subregions, the CA1 marker gene Wfs1 was the most highly expressed by more than tenfold and strong expression of Wfs1 is described in Dong et al (2009) as uniquely defining the dorsal domain of CA1. Expression of markers of CA1 within the hippocampus was significantly higher than expression of markers of other subregions and expression of dCA1 markers was significantly higher than markers of vCA1. Expression of vCA1 markers was negligible suggesting that the virus did not spread to vCA1, in accordance with the very weak fluorescent signal in this region. Markers of dCA1, except for Wfs1, were expressed at a similar level to markers of astrocytes and PVINs. Combined with the other results, this suggests that expression of these CA1 markers is relatively weak compared to the expression of key neural cell type markers. Markers of other hippocampal subregions were weakly expressed in the samples. There was evidence of potential enrichment of genes expressed in CA2, the area most likely to be dissected alongside CA1, compared to CA3 and DG. The findings may also be affected by differences in gene expression between species as the cited analyses of domain-specific expression were conducted using mice.

In summary, transcripts from CA1 pyramidal neurons were found to be the most abundant in the samples, with PVINs and astrocytes transduced tenfold less. The results also suggest that virus expression within CA1 was specific to the targeted dorsal domain.

3.4.5 Conclusions

A previous report of higher Fos expression in CA1 following paired viewing of novel compared to familiar arrangements was validated. This behavioural procedure was combined with the extraction of ribosome-bound mRNA using the TRAP method and with RNA sequencing to profile the translatome of CA1 neurons during associative recognition memory formation.

Purified TRAP-generated RNA was extracted from rodents following paired viewing and a suitable method of RNA amplification was identified in order to generate sufficient RNA for sequencing. Following quality control of sequencing data, samples with low RIN values were excluded and data was pre-processed to remove library preparation artefacts. The methods could be improved in future experiments by purifying reagents to reduce contamination and using an RNase-free hood to reduce RNA degradation.

Sequencing data suitable for downstream differential expression analyses was obtained and analysis of marker gene expression in the dataset indicated that the TRAP samples

were enriched for CA1 neuron markers. These data are further analysed in the following chapter.

3.5 Supplemental Figures

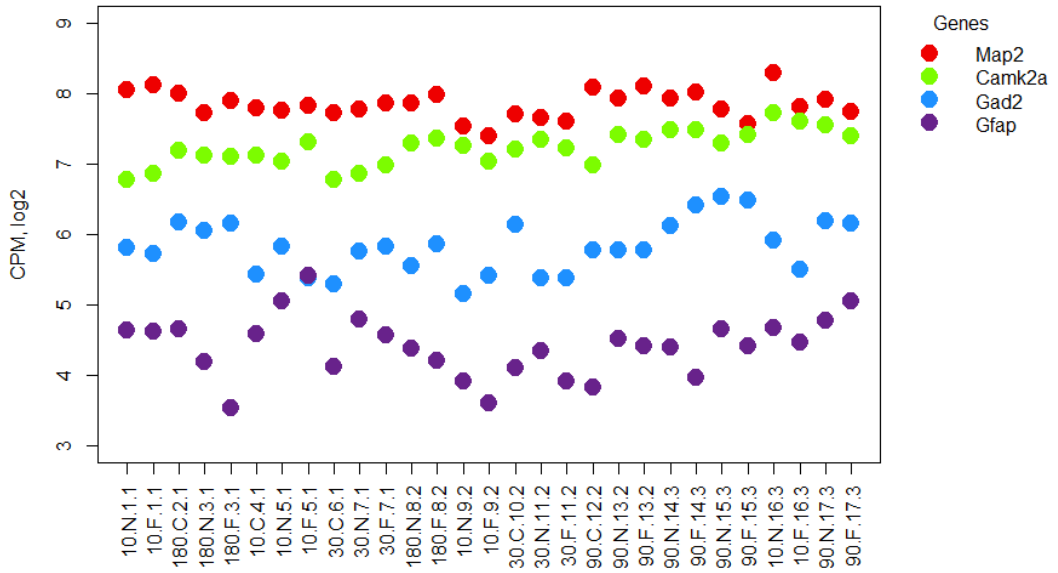


Figure 3.26: Expression of key neural cell type markers in individual samples. Log_2 CPMs of markers of neurons (Map2), excitatory neurons (Camk2a), interneurons (Gad2) and glia (Gfap) in sequenced samples.

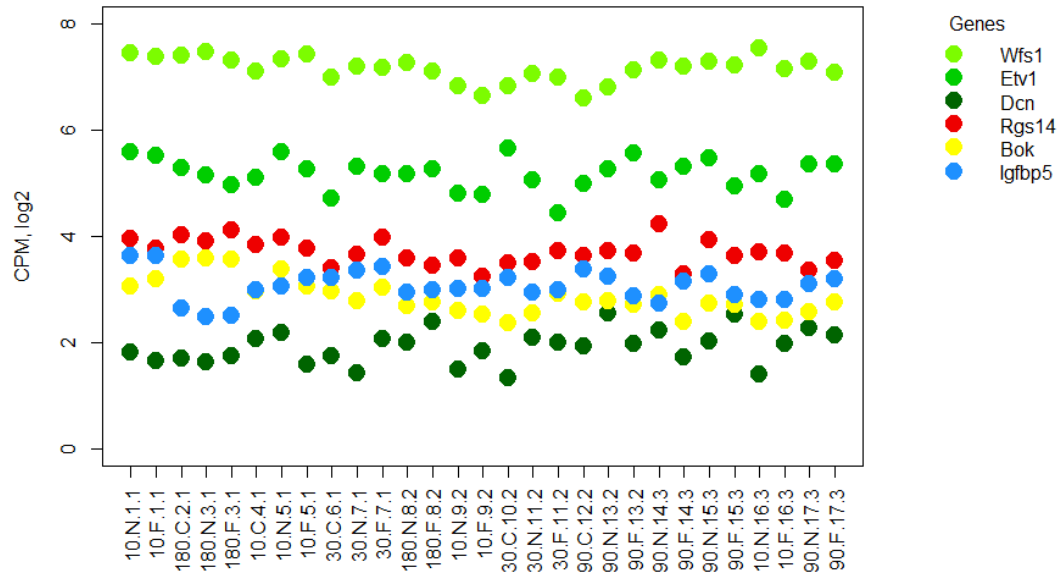


Figure 3.27: Expression of selected markers of hippocampal subregions and CA1 domains in individual samples. Log_2 CPMs of markers of CA1 (Wfs1, Etv1), CA1d (Wfs1), CA1v (Dcn), CA2 (Rgs14), CA3 (Bok) and DG (Igfbp5) in sequenced samples.

Chapter 4 Profiling the CA1 translatome during associative recognition memory formation

4.1 Introduction

This chapter presents the analysis of gene expression matrices described in Chapter 3, with a focus on how to conceptualise, analyse, and validate findings from this complex multi-factor dataset. There is no one accepted approach to analysis and comparison studies have reported that the best approach is dependent on the dataset and the study objectives (Rapaport et al, 2013; Conesa et al, 2016; Lin et al, 2016; Mohorianu et al, 2017).

In the present analysis, comparison of gene expression between conditions is of interest as opposed to comparison of the expression of different genes within samples and for this purpose, count data is more suitable for downstream analyses than methods that normalise expression by gene length. Furthermore, as this work is primarily exploratory, some false positives are tolerated in order to reduce the type II error rate and increase power to detect differential gene expression. The following introductory sections discuss approaches to modelling differential expression in NGS datasets and interpretation of the findings in a biological context.

4.1.1 Statistical modelling of gene expression data

4.1.1.1 Approaches to normalisation

Gene expression measurements are made relative to other genes in the sample and are subject to unwanted variation from multiple sources in addition to biological variation (Robinson & Oshlack, 2010). Hence the normalisation of NGS data is essential to allow different datasets to be compared. The normalisation of gene expression data is challenging because the proportion of reads in a sample, for genes with stable absolute expression, is affected by differences in global expression between samples and by sample processing (Robinson & Oshlack, 2010; Conesa et al, 2016). Normalisation by library size is therefore inadequate for differential expression analysis.

4.1.1.1.1 Global adjustment methods

Broadly, normalisation approaches involve adjusting values to a common scale or aligning sample probability distributions. Two of the simplest methods to normalise RNAseq count data, alignment of the median or upper quartile (UQ) of each sample (Dillies et al, 2013), are examples of the former approach. Compared to total count normalisation, these methods prevent normalised expression values from being skewed by highly expressed genes and highly variable genes (Bullard et al, 2010). In alternative distribution-based approaches to normalisation, such as the trimmed mean of m-values (TMM; Robinson & Oshlack, 2010) and DESeq (Anders & Huber, 2010) methods, a scale factor is estimated for each library in a dataset based on the assumption that most genes are not differentially expressed between samples. For experiments that may

involve global differences in expression, this assumption does not hold, but this is unlikely to be of concern here. However, it is possible that the distribution of gene expression in the translatome is different to the transcriptome, for which these methods of analysis were developed.

As a distribution-free alternative, a recent paper used a normalisation method based on subsampling without replacement (Mohorianu et al, 2017). Another alternative is to use multiple rounds of iterative normalisation so that normalisation is conducted without inclusion of potential DEGs, as implemented in the R package Tag count comparison (TCC; Sun et al, 2013). This process can be used with many different combinations of normalisation and differential expression methods.

Normalisation is also affected by genes with low expression, for which measurements are less reliable, and so these genes are usually filtered prior to normalisation (Oshlack & Wakefield, 2009), though it has also been suggested that the opposing approach is less biased (Lin et al, 2016). As sequencing depth in the present study was variable and relatively low after the pre-processing steps described in Chapter 3, expression of low-count genes was unlikely to be comparable across samples and so these genes were removed prior to normalisation.

4.1.1.1.2 Accounting for unwanted variation

Batches of samples prepared and sequenced at different times are affected by unwanted variation which can have major effects on the results of sequencing experiments (Taub et al, 2010; 't Hoen et al, 2013) and increase both type 1 and type 2 error (Gagnon-Bartsch & Speed, 2012). Ideally, all samples in an experiment are prepared and sequenced in parallel or samples from different conditions are counterbalanced across batches, but this may not be achieved in practice. Batch effects may be correlated with experimental factors and can also affect correlations between genes, such that two genes may be positively correlated in one batch and negatively correlated in another (Leek et al, 2010). Batch effects are not accounted for by the normalisation methods discussed above, but several methods of correction have been developed and are particularly important where effects of interest are likely to be small and/or highly variable, as is often the case in behavioural studies of learning and memory (Leek et al, 2010; Peixoto et al, 2015b).

A useful method for data exploration is principal components analysis (PCA), which enables multi-dimensional data to be visualised in two-dimensional space by identifying the linear combinations of variables that maximise variance, the principal components (PCs). PCA can be used to clarify the structure of a complex dataset and determine the main sources of variation, such as the presence of batch effects. In gene expression datasets with relatively subtle manipulations, the first PC often corresponds to sequencing batch (Leek et al, 2010; Peixoto et al, 2015b). PCA can also be suggestive of factors that may be removed to reduce model complexity in further analyses (Bradlow, 2002).

Batch effects are caused by a variety of factors related to both the preparation and sequencing of samples and though correlated with batch, are not explained by batch

alone (Leek et al, 2010). The most effective methods of correction for batch effects are therefore based on the detection of latent variables that represent unwanted sources of variation (Li et al, 2014), such as the methods implemented by RUVseq (Remove Unwanted Variation from RNAseq data; Risso et al, 2014). Weights that represent factors of unwanted variation in each sample can then be incorporated into differential expression analysis (Leek & Storey, 2007). If instead gene expression datasets are adjusted for batch effects prior to analysis, degrees of freedom are reduced (which may result in inaccurate modelling of test statistics) and group differences may be exaggerated (Nygaard et al, 2016). Including factors of unwanted variation in the experimental design, on the other hand, can account for batch effects that are correlated with experimental effects and enable modelling of greater effects of unwanted variation on some genes than on others (Leek et al, 2010; Risso et al, 2014; Nygaard et al, 2016).

In RUVseq, adapted from the RUV methods for microarrays (Gagnon-Bartsch & Speed, 2012), factor analysis is used to estimate factors of unwanted variation (FUVs) using negative control genes, biological replicates, or the residuals from a first-pass generalised linear model (GLM) regression on the co-variates of interest (Risso et al, 2014). The latter two approaches assume that unwanted variation is not correlated with the variables of interest, which may not be the case in the present dataset, and so FUVs were estimated using negative control genes. Empirical negative control genes were determined from the data as a reliable set of a priori negative controls was not available and effects of condition were expected to be relatively subtle, so it could be assumed that genes with high p-values on a first-pass differential expression analysis would not be affected by the experimental conditions.

4.1.1.2 Differential expression

There are many approaches to the identification of differentially expressed genes (DEGs) between conditions in RNAseq data. Due to the differences in outcome between tools, differential expression (DE) analysis often involves comparison of multiple pipelines (e.g. Lin et al, 2016; Mohorianu et al, 2017).

4.1.1.2.1 Approaches to differential expression analysis

Gene expression data can be modelled using probability distributions, which describe the range of values taken by different types of data as a function of parameters such as mean and variance. The choice of distribution is influenced by whether the data is discrete (e.g. raw counts) or continuous (e.g. RPKM), data symmetry, limits on the possible range of values, and the likelihood of observing extreme values.

Discrete probability distributions using gene counts are often used to model RNAseq data (Conesa et al, 2016). Individual RNAseq samples and technical replicates follow a Poisson distribution (Marioni et al, 2008), which is parameterised by a ‘rate’ corresponding to the average number of events in an interval and has variance equal to the mean. However, modelling the additional variability that exists between biological replicates requires extensions of the Poisson distribution that can incorporate over-dispersion, such as the negative binomial (NB) and beta binomial distributions (Robinson

& Smyth, 2007; Hardcastle & Kelly, 2013). The DE tools edgeR (Robinson et al, 2010), baySeq (Hardcastle & Kelly, 2010), and DESeq (Anders & Huber, 2010) are all based on the NB distribution, which models the number of trials needed to reach a fixed number of successes. The NB distribution has an additional parameter compared to the Poisson and is more positively skewed with similar parameters. These methods are described in more detail in the following section.

Differential expression in RNAseq data can also be modelled using normal distributions with Limma-voom (Law et al, 2014). Limma was developed for microarrays and is adapted for RNAseq using a method termed 'variance modelling at the observational level' (voom). Although RNAseq data is not normally distributed, normal distributions have the major advantage of being more analytically tractable than discrete probability distributions and so in the Limma-voom method, log-normalised counts are modelled and precision weights are assigned to account for the mean-variance trend, which is fitted using an empirical Bayesian method (Law et al, 2014). Alternatively, non-parametric approaches can be used that make minimal distributional assumptions and estimate empirical distributions directly from the data, such as NOIseq (Tarazona et al, 2011) and SAMseq (Li & Tibshirani, 2013). However, these methods usually require relatively large numbers of biological replicates (e.g. N>10; Conesa et al, 2016).

There are few DE tools that enable analysis of multiple conditions and factors simultaneously and some are limited to specific types of design. In edgeR (McCarthy et al, 2012) and DEseq2 (Love et al, 2014), arbitrarily complex experimental designs can be analysed using generalised linear models, an extension of general linear modelling for distributions from the exponential family, including the Poisson and NB. This approach enables incorporation of interaction terms, batch effects, and pairwise comparisons and is described in detail below.

4.1.1.2.2 Statistical theory underlying parametric models for differential expression

EdgeR and DESeq2 apply broadly similar approaches to model differential expression in multi-factor RNAseq datasets using GLMs (McCarthy et al, 2012). GLMs extend ordinary linear regression models (of the general form $Y = mX + c$) to response variables with residuals that are not normally distributed by using a link function to enable a log-linear model of the mean-variance relationship to be fitted. To model RNAseq data, observed counts for each gene in each sample are modelled as a function of the variance of counts (referred to as dispersion; Robinson & Smyth, 2008) and of each of the factors included in the design with a sample-specific offset, which is determined by the global method of normalisation (see section 4.1.1.1.1). As different genes may have different mean-variance relationships, yet the number of replicates in RNAseq experiments is usually too low to estimate these relationships accurately for individual genes, gene-specific estimates of the dispersion parameter of the NB distribution can be moderated towards global dispersion to permit information sharing across genes (Robinson & Smyth, 2007). After fitting the dispersion parameters and GLMs, statistical tests for differential expression between conditions are conducted and p-values are corrected for multiple testing. These steps are described in more detail below.

Global dispersion in edgeR can be estimated using the variance parameter from the NB distribution for all genes (common dispersion; Robinson & Smyth, 2008) or by fitting a trended dispersion to account for heteroscedasticity, i.e. differences in variance that are dependent on gene abundance. For example, low-count genes often have higher variance due to measurement error.

Gene-specific dispersion estimates are fitted to account for genes with variance above or below the common or trended variance so that genes with more consistent behaviour across replicates are ranked more highly in the DE analysis (Robinson & Smyth, 2007). In both edgeR and DESeq2, empirical Bayesian methods are used to moderate gene-specific estimates of dispersion towards global dispersion. Bayes' Theorem accurately specifies the probability of one event given another (Equation 2; Equation 3). The *posterior* probability, i.e. the probability of 'A' given that condition 'B' is true, is calculated by multiplying the *likelihood* of 'B' given that 'A' is true by the probability of 'A' (the probability *prior* to knowing 'B') and dividing by the probability of 'B' (the *evidence*). In other words, the likelihood is the probability of obtaining the observed data given a model, the prior is the probability of the model, and the evidence is the probability of the data.

Equation 2

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Equation 3

$$Posterior\ probability = \frac{Likelihood * Prior\ Probability}{Evidence}$$

In edgeR, the Cox-Reid profile-adjusted likelihood empirical Bayesian method is used to determine the degree of moderation towards the global dispersion, dependent on the precision of the gene-specific estimates and the prior degrees of freedom in the data (McCarthy et al, 2012). In DESeq2, gene-specific dispersions are estimated first by determining the values that maximise likelihood of the observed counts and a curve is fitted to the dispersion estimates. Gene-specific estimates are then shrunk towards the fitted values using an empirical Bayesian method (Love et al, 2014). These applications of empirical Bayesian methods can be contrasted with Bayseq (Hardcastle & Kelly, 2010), in which null and differential expression models are defined for each gene and used to estimate the posterior probability of differential expression.

Next, NB GLMs are fitted using the moderated dispersion estimates for each gene. In both edgeR and DESeq2, the likelihood ratio test (LRT) can then be used to perform an ANOVA-like test for differences between multiple conditions simultaneously. In edgeR, the quasi-likelihood (QL) method was recently introduced to reflect the uncertainty in gene-specific dispersion estimates by incorporating both overall biological variability and gene-specific variability into the GLM, thereby providing greater control of type 1 error (Lun et al, 2016). This method uses the QL F-test for differential expression.

To deal with outliers in edgeR, weights can be assigned to more variable genes to reduce their influence on GLM fitting (Zhou et al, 2014). Alternatively, in DESeq2, outlying counts are identified using Cook's distance and these genes are removed (or outliers are replaced with imputed values if there are >6 replicates; Love et al, 2014). To correct for the false discovery rate (FDR) caused by multiple testing, the above methods all use the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). To improve power to detect differential expression, DEseq2 can additionally perform independent filtering of low-count genes to achieve a specified FDR.

EdgeR and DEseq2 have generally been found to have similar performance statistics and perform well compared to other methods in comparison studies (Soneson & DeLorenzi, 2013; Love et al, 2014; Zhou et al, 2014; Lin et al, 2016), though these methods may identify different genes (Soneson & DeLorenzi, 2013). Both tools were applied to analyse the present dataset and alternative methods were compared.

4.1.2 Biological insight from RNAseq data

Genome-wide expression profiling studies provide rich datasets to which a wide variety of methods can be applied to gain biological insight, including integrating information from databases, building networks, and machine learning. In this section, methods used in this thesis to interpret differentially expressed genes will be outlined, including approaches to gene ontology analysis (Ashburner et al, 2000), identifying protein-protein interactions, and clustering of co-expressed genes.

4.1.2.1 Finding enriched functional groups

To interpret the large numbers of genes usually identified in genome-wide studies, databases of information about gene function are invaluable and are often used in conjunction with statistical tests for the enrichment of genes annotated to specific categories, which may be conducted using distinct lists of genes, gene ranks, or expression matrices (Goeman & Buhlmann, 2007).

4.1.2.1.1 Gene Ontology

The Gene Ontology (GO; Ashburner et al, 2000; GO Consortium, 2017) system of functional annotation is frequently used in RNAseq data analysis. GO terms are separated into three ontologies: molecular function (MF), biological process (BP), and cellular compartment (CC). MF terms describe direct molecular actions such as "binding", while BP terms relate to higher-level processes and biological systems such as "learning and memory" or "regulation of synaptic plasticity", and CC terms are cellular locations or structures such as "post-synaptic density". GO terms within each ontology are organised in a directed acyclic graph with specific terms nested within more general ones, providing a structure to categorise gene function at multiple levels. Relationships between terms are also specified, for example "is_a" or "regulates".

The results of GO enrichment analyses can be difficult to interpret as many redundant terms are returned and similar sets of genes are often involved in different functions in different cell types and systems, resulting in enrichment of terms that are not relevant to the system under study. Many approaches have been devised to account for the

relationships between GO terms. In enrichment analysis, the graph structure of ontologies can be exploited to reduce redundancy between terms, using different algorithms to favour broader (parent) terms (e.g. Alexa et al, 2006) or more specific (child) terms (e.g. Grossmann et al, 2007) or to weight the two (e.g. Alexa et al, 2006). The results of these analyses can be visualised within the GO graph structure (Zhang et al, 2005; Alexa et al, 2006). Another approach has been the curation of GO Slim vocabularies, containing a reduced set of annotations (Camon et al, 2004).

However, to assess the functional relationships between GO terms that are not connected by parent-child relationships within the GO graph and those from different ontologies, different approaches are necessary. For example, in DAVID (Dennis et al, 2003), a popular online tool for GO analysis, annotations and genes can be grouped into "biological modules" by fuzzy clustering of their co-occurrences across multiple sources of information (Huang et al, 2007). The relationships between sets of annotations can also be based on the semantic similarity of terms (Yu et al, 2010) or on gene set overlap between terms (Merico et al, 2010). Measures of similarity can further be used to visualise the relationships between GO terms enriched in a dataset as a map and in this way, compare different conditions (Rhodes et al, 2007; Merico et al, 2010). A major limitation of using GO to interpret findings is that only already-known functions of genes can be considered and well-studied functions and genes will be represented disproportionately. However, gene functions can be identified more consistently and with more confidence than individual genes (Merico et al, 2010) and the relationships between GO terms are more statistically stable than relationships between genes (Zhang, 2015). Thus, the detection of enriched GO terms can in some respects be considered more reliable than the detection of DEGs.

4.1.2.1.2 Additional databases for interpretation of differentially expressed genes

In addition to GO, alternative databases that are useful for the functional analysis of large gene sets include the PANTHER protein classes (Thomas et al, 2003), which are related to GO molecular function terms, but provide a distinct set of annotations with somewhat broader descriptions of protein functions, and the Kyoto encyclopedia of genes and genomes (KEGG; Kanehisa & Goto, 2000) pathway database, which describes metabolic and signalling pathways, as well as sets of genes linked to disease or that are targeted by certain drugs.

For evidence of interactions between individual genes, STRING (von Mering et al, 2003) is an informative database of protein-protein interactions, which contains evidence from a variety of direct and indirect sources, including published experiments, co-expression, text-mining of abstracts to find genes mentioned together, and association of genes in other databases. Interaction scores between zero and one are assigned for each source and for all sources combined according to the strength of evidence for an interaction in the specified species using evidence from all organisms. This information can be used to determine whether interactions within a dataset are enriched compared to the expected number of interactions and to build and analyse protein-protein interaction networks, enabling identification of clusters of interacting genes and potential hub proteins that are central to the network (He & Zhang, 2006).

4.1.2.2 Clustering

Gene expression can be viewed as the output of the co-ordinated activity of multiple regulators (D'haeseleer et al, 2000; Mahajan & Mande, 2015). Many studies have shown that underlying patterns in gene expression datasets, such as groups of co-regulated genes, can be identified using clustering algorithms (Eisen et al, 1998; Jiang et al, 2004; Yeung et al, 2004; Li & Bushel, 2016) and may correspond to biological functions and/or the activity of particular transcription factors or signalling pathways (Tornow & Mewes, 2003; Di Giovanni et al, 2004; Iacono et al, 2013; Liu & Si, 2014). Clustering can also reveal differentially expressed patterns that are missed by analyses that focus on individual DEGs (Li & Bushel, 2016). Common approaches to clustering will be outlined here.

Clustering algorithms are methods for unsupervised machine learning, which enable computers to learn to make predictions from data that improve with experience without prior knowledge of the data structure. This is achieved by grouping items (e.g. genes or samples) with similar features. Clustering contrasts with supervised machine learning algorithms, which utilise labelled training data to identify features that permit future classification of unlabelled data, such as identifying gene signatures corresponding to disease phenotypes (Wei et al, 2009). Many approaches to clustering gene expression data have been proposed and each method emphasises different features of the data and will group genes differently, which may be useful for different applications (D'haeseleer et al, 2000; Liu & Si, 2014).

Approaches to clustering include hierarchical methods, in which clusters are nested, and partitioning methods, in which all clusters are on the same level. Hierarchical clustering algorithms (e.g. Alon et al, 1999) can be agglomerative, beginning with each object as its own cluster and proceeding by repeated merging of the two closest clusters, or divisive, beginning with all elements in one cluster which is then repeatedly divided until all elements are separated. In contrast, k-means clustering (e.g. Iacono et al, 2013; TimesVector, Jung et al, 2017) requires a number of clusters (k) to be specified and non-overlapping clusters are then formed by defining k random starting points, iteratively assigning objects to the nearest cluster, recalculating cluster centres, and reassigning objects until the algorithm converges. This process is repeated with different initialisations to optimise the clustering configuration. The number of clusters can be optimised separately, for example using the gap statistic (Tibshirani et al, 2001). Model-based methods (e.g. DGEclust, Vavoulis et al, 2015) are another form of partitioning and represent data as a random sample from a mixture of probability distributions, each of which corresponds to a distinct cluster. Whereas hierarchical and k-means clustering are best suited to normally distributed data and may require data transformation (Liu & Si, 2014), model-based methods can be used to fit discrete distributions such as the NB, which is often used to model count data (see section 4.1.1.2.1). Model-based methods, as well as other algorithms such as fuzzy c-means, also enable soft clustering, in which items can belong to more than one cluster or cluster membership is probabilistic. Most forms of clustering require a distance metric to quantify similarity (or dissimilarity) between data points, the most common being Euclidean (geometric) distance and

Pearson correlation-based (Liu & Si, 2014). To cluster time-series data, it may be necessary to account for correlations between adjacent timepoints (Spies & Ciaudo, 2015) and this can be achieved by using cosine dissimilarity as a distance metric, as in the algorithm TimesVector (Jung et al, 2017).

The major advantage of clustering is that it enables dimensionality reduction, the grouping of overlapping features to represent a dataset using a smaller number of factors. If a dataset has meaningful clusters, the distance between groups will explain a large proportion of the overall variability in the data. The value of clustering can also be assessed by biological plausibility in combination with functional enrichment analyses, as discussed above.

Clustering was applied in this thesis with the aim of organising DEGs according to their expression profiles and thus providing a summary of the broad changes in expression over time following exposure to different stimulus conditions. As the timepoints in the present study were unevenly spaced and from independent samples and as regulation of the translatome has been shown to be highly dynamic (Cho et al, 2015b; Chen et al, 2017), the utility of accounting for time in this analysis was limited. Instead, a k-means approach to clustering was applied, similar to the methods used previously by Dijkmans et al (2009) and Iacono et al (2013) to analyse highly non-stationary time-course data. Clustering was conducted separately for each condition due to the dependence between novel and familiar samples from the same timepoint and to reflect that the same genes may be differently regulated by overlapping pathways under different conditions (Kotaleski & Blackwell, 2010). DEGs were selected as features of interest prior to clustering to reduce the effects of random or task-irrelevant fluctuations in gene expression on clustering (Liu & Si, 2014).

4.1.3 Present work

Genome-wide translatome profiling data from rat CA1 neurons was obtained at a series of timepoints following paired viewing of novel and familiar visuospatial stimuli or a no-image control condition, as described in Chapter 3. The primary aim of the study was to use differential expression analysis to identify genes that may be involved in the formation and consolidation of associative recognition memory. The expression of IEGs including Fos and Arc was used as a positive control for gene expression changes associated with the early stages of memory consolidation. Based on previous studies of IEG expression following exposure to novel stimuli (Guzowski et al, 1999; Ramírez-Amaya et al, 2005; Vazdarjanova et al, 2006) and the finding of greater Fos expression in CA1 in the novel condition following paired viewing of arrangements (Wan et al, 1999), it was hypothesised that the strongest induction of IEGs would be observed in the novel condition. The no-image condition was used to control for expression changes due to other aspects of the behavioural paradigm such as stress or motor activity, as has been observed in previous experiments using behaviourally-matched control conditions (Bertaina-Anglade et al, 2000; Cavallaro et al, 2002; Shires & Aggleton, 2008).

The first stage of the analysis, presented in section 4.3.1, was the optimisation of the pipeline for differential expression analysis. Firstly, approaches to filtering and

normalisation were identified that would reduce noise and control for unwanted variation due to batch effects. Then, methods for differential expression testing based on NB GLMs were compared to find the most suitable model for the variance structure of the data, as discussed in section 4.1.1.2. Preliminary analyses of gene ontology, expression time courses, and co-expression of functionally related genes were then conducted to determine the scope of the analysis based on the strength and biological plausibility of patterns in the data. Based on these results, downstream analyses were focused on the DEGs identified in an ANOVA-like test for differential expression across all conditions using the optimised pipeline.

In section 4.3.2, upregulated expression of IEGs at early timepoints following paired viewing was confirmed as a positive control.

In section 4.3.3, a range of analyses were conducted on the DEGs identified by the pipeline optimised in section 4.3.1 to study the molecular processes involved in associative recognition memory. The types, functions, and interactions of DEGs were investigated using databases and k-means clustering was used to explore the relationship between the timing of expression changes and known protein functions.

4.2 Methods

Analyses were conducted on the gene expression count tables generated in Chapter 3. The number of samples in each condition in the final analysis is shown in Table 4.1. Analyses were conducted in R 3.4.1 using packages available through Bioconductor (Gentleman et al, 2004) except where otherwise stated.

Table 4.1: Number of samples in each condition

		Experimental condition		
		Familiar	***Novel***	***No-image control***
Timepoint (minutes)	***10***	4	4	1
	30	2	2	1
	90	4	4	1
	180	2	2	1

4.2.1 Detection of differentially expressed genes

4.2.1.1 Normalisation

Genes were filtered to retain only those with CPM>2 in a minimum of two samples (N=13821). The threshold for the removal of low-count genes was determined by CPM as opposed to raw counts to prevent a bias of library size on filtering. UQ normalisation was applied to the remaining genes to account for differences in library size. PCA of the normalised counts was conducted using the EDASeq package (Risso et al, 2011).

Factors of unwanted variation between negative control genes were found using RUVseq (Risso et al, 2014). Negative control genes (N=6263) were empirically defined as those with p>0.5 on a first-pass differential expression analysis of all conditions using the edgeR LRT robust (LRTr; Robinson et al, 2010; McCarthy et al, 2012; Zhou et al, 2014) pipeline as described below.

4.2.1.2 Differential expression tests

Differential expression analysis was conducted using edgeR (Robinson et al, 2010) and DESeq2 (Love et al, 2014) to compare different algorithms based on NB GLMs. In edgeR, DE analyses were conducted using the chi-square approximation to the LRT (McCarthy et al, 2012) and the QL F-test (Lun et al, 2016), both with and without robust dispersion estimation (Zhou et al, 2014). In the design matrix, image condition and timepoint were combined into one experimental factor, as opposed to modelling two main factors and an interaction, to account for the possibility that different effects on gene expression would be present for each combination of image condition and timepoint (McCarthy et al, 2012). For no-image control samples, timepoint was not specified in the design matrix so that control samples were treated as biological replicates for DE analysis to increase degrees of freedom to estimate the dispersion parameters. The first two factors of unwanted variation detected using RUVg as described above were included in the design matrix to account for batch effects.

Coefficients for all experimental conditions were specified to perform ANOVA-like tests for differences between all conditions simultaneously, using the no-image control condition (all timepoints) as the intercept (McCarthy et al, 2012). Genes with false discovery rate (FDR) values below the commonly used threshold of 0.1 (e.g. Chen et al, 2017) were selected for further analysis. DEGs detected by different pipelines were compared using the R package VennDiagram (Chen & Boutros, 2011). Based on the results of the comparison, the LRTr method was used for further analyses. Using the LRTr pipeline, contrasts of novel and familiar conditions were conducted at each timepoint.

4.2.2 Functional interpretation

4.2.2.1 Hierarchical clustering

Agglomerative hierarchical clustering was conducted on DEGs using the complete-linkage method. One minus Pearson-correlation was used as the dissimilarity index and was computed on the matrix of $\log_2$ fold changes (FCs) estimated by DE analysis for selected genes in each condition compared to pooled controls. The R package gplots (Warnes et al, 2016) was used to plot heatmaps and dendrograms. Log_2 FCs for each gene were standardised by mean-centring and scaling to unit variance for plotting.

4.2.2.2 Gene ontology

GO (Ashburner et al, 2000; GO Consortium, 2017) enrichment analyses were conducted using the R package topGO (Alexa & Rahnenführer, 2010) to implement Fisher's exact test. The elimination algorithm introduced by Alexa et al (2006) was used to reduce redundancy between significant terms. GO terms with fewer than five annotated genes were not analysed. Enrichment of GO terms among selected genes was compared to a background of all expressed genes (N=13821).

PANTHER online tools (Mi et al, 2010) were used to count DEGs annotated to PANTHER protein classes (Thomas et al, 2003).

4.2.2.3 K-means clustering

K-means clustering was performed on DEGs detected using the edgeR LRTr method as described above. Prior to clustering, counts were adjusted to remove two factors of unwanted variation using RUVseq. Mean gene expression values for each condition were then calculated and standardised by mean-centring and scaling to unit variance.

K-means clustering was performed using the Hartigan & Wong (1979) algorithm, with 25 random starting positions and 30 iterations. The seed for random number generation was set to 123. To determine the optimum number of clusters, the gap statistic (Tibshirani et al, 2001), which refers to the gap between the total within-cluster sum-of-squared errors (SSE) for the data compared to null reference data, was calculated for values of k (number of clusters) between one and ten using the R package cluster (Maechler et al, 2017). To calculate the gap statistic, B=500 bootstrapped reference samples were generated with null distributions and values within the range of each variable using Monte Carlo simulations. For each value of k, the mean log within-SSE was calculated for bootstrapped samples and subtracted from the log within-cluster SSE to

obtain the gap statistic. The value of k used for subsequent clustering was determined by the first maximum of the gap statistic. All optimisation algortihms converged within thirty iterations.

4.2.2.4 Mapping enriched GO terms

Enriched GO terms in DEG clusters in novel and familiar conditions were imported into Cytoscape v3.5.1 (Shannon et al, 2003). To quantify gene set overlap between GO terms, similarity coefficients were calculated using the Cytoscape application EnrichmentMap (Merico et al, 2010). From these data, a network of significant GO terms was generated in which edges represented a similarity coefficient > 0.3. Overly general GO terms with more than 5000 annotated genes were removed. An enrichment significance threshold of $p < 0.025$ was chosen to include a sufficient number of GO terms to visualise the relationships between gene functions and experimental conditions. Visual properties of the network were mapped to factors of interest such as enrichment p-value. The "prefuse" force-directed layout (Heer et al, 2005) was applied, which is determined by a simulation of interacting forces in which nodes repel one another and edges act as springs, with edge length weighted by the similarity coefficient. Clusters were identified and labelled manually based on inspection of the network topology and GO terms and the network layout was adjusted to reduce overlap and emphasise groups of connected nodes related to similar functions.

4.2.2.5 Protein-protein interaction networks

DEG lists were submitted to STRING (von Mering et al, 2003) to identify protein-protein interactions. To map visual properties of the network to experimental variables and differential expression analysis results, STRING data was imported into Cytoscape. Betweenness centrality scores (Brandes, 2001) were calculated in Cytoscape to identify candidate 'hub' proteins.

4.3 Results

4.3.1 Optimisation of differential expression analysis

4.3.1.1 Normalisation

For edgeR DE analysis, genes with CPM>2 in ≥2 samples were retained (N=13821). For DESeq2 DE analysis, which filters low-expressed genes internally, genes with a minimum of 10 raw counts across all samples were retained (N=22029).Figure 4.1 shows the distribution of relative log_2 expression ratios (RLEs) of filtered genes in each sample before and after UQ normalisation. Figure 4.1A shows that sample means are variable prior to normalisation. Overall differences in distribution can be seen between batches but are not consistent across all samples within a batch. For example, most but not all samples in batch 1 have mean RLE below zero, whereas most but not all samples in batch 3 have mean RLE greater than zero. Figure 4.1B shows that UQ normalisation aligns the centre of all sample distributions but does not align the tails of distributions. Batch 3 samples have shorter tails than most of the samples in batches 1 and 2 suggesting that unwanted variation related to batch is present.

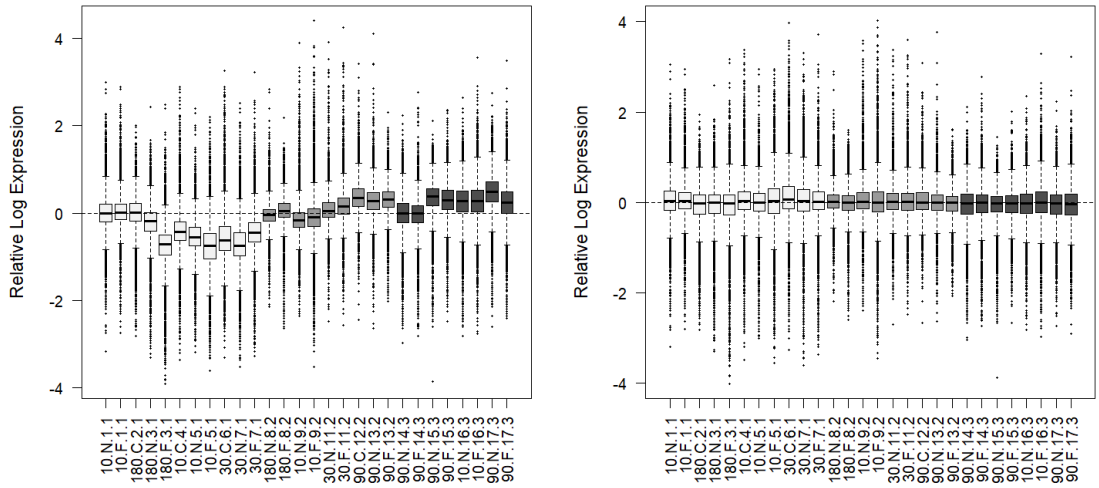


Figure 4.1: Distribution of relative log_2 expression values (RLEs) before and after upper quartile (UQ) normalisation. An RLE of zero corresponds to the median log_2 expression across all samples. Sample labels along the x-axis indicate timepoint, image condition, group of animals, and sequencing batch. Colour indicates batch. A: Prior to normalisation, but after filtering low-expressed genes, samples have variable mean RLEs and distributions. Most samples in batch 1 have means below zero, whereas most samples in batch 3 have means above zero. B: After UQ normalisation, sample means are aligned but overall distributions remain variable. (N=13821 genes)

Figure 4.2A shows PCA after normalisation and filtering. There was some separation between samples from early (10-30 minutes) and late (90-180 minutes) timepoints within batches by PC1. However, the first two PCs of variability between samples were both affected by sequencing batch and neither PC alone separated the three batches. The first two PCs accounted for a relatively low proportion of overall variance (14.45% and 11.13% respectively), presumably due to the large number of factors involved. The RUVg method (Risso et al, 2014) was used to account for batch effects and other sources of unwanted variation, using factor analysis to find weights for each sample to adjust for

unwanted variation in the expression of negative control genes (N=6433 genes with p>0.5 in a first-pass DE analysis using the edgeR LRTr test (Robinson et al, 2010; McCarthy et al, 2012; Zhou et al, 2014)). Figure 4.2B shows PCA of counts adjusted to remove two factors of unwanted variation. Sequencing batch effects have been removed or greatly reduced, but samples do not cluster by experimental condition in two dimensions. However, there is partial separation between samples from early and late timepoints along PC1, which accounts for 16.03% of variance. PC2 accounts for only 7.29%. These results suggest that effects of condition are relatively subtle compared to overall variability.

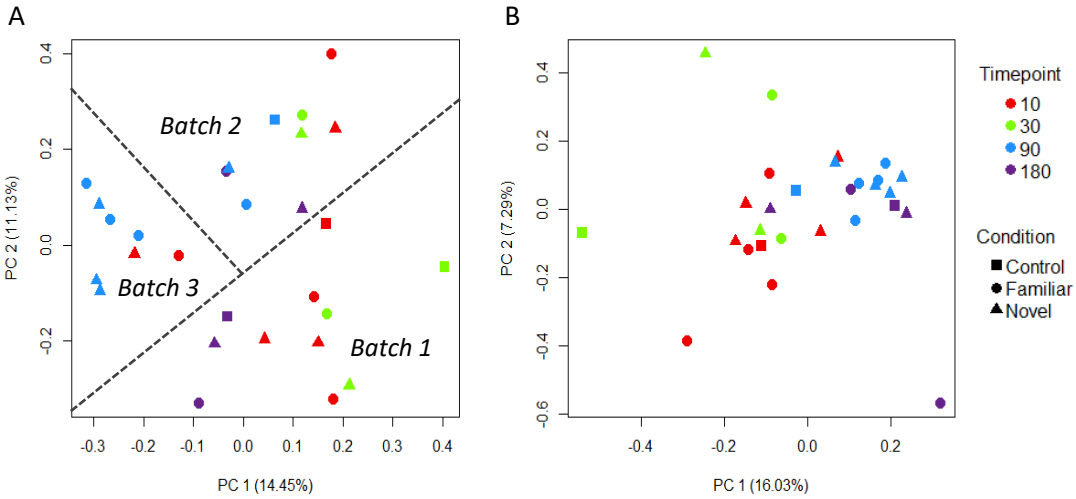


Figure 4.2: Principal components analysis (PCA) of PV-TRAP samples before and after RUVg normalisation. Timepoint is indicated by colour and stimulus condition is indicated by shape. A: Separation between samples after upper quartile normalisation is driven by batch. B: Two factors of unwanted variation were removed using RUVg. Batch effects are reduced and clustering of samples from early vs late timepoints is improved.

4.3.1.2 Methods of differential expression

Alternative DE pipelines using the R packages edgeR (Robinson et al, 2010) and DESeq2 (Love et al, 2014) were compared to optimise the specification of the design matrix and the choice of DE test, guided by examination of the p-value distribution (Hu et al, 2010a), the number of DEGs identified in each test, and the overlap in identified DEGs between tests. Gene expression was compared across all conditions, using the no-image control samples from all timepoints as a baseline. The methods of DE testing compared in this section are the edgeR LRT (McCarthy et al, 2012) and QL F-test (Lun et al, 2016), both with and without robust dispersion estimation (Zhou et al, 2014), and the DESeq2 likelihood ratio test (Love et al, 2014).

Alternative design matrices were compared to determine the most reliable model for unwanted variation. Figure 4.3Ai-iii shows the distribution of p-values for DE analyses using the edgeR LRTr method. Without accounting for batch, there is an inflation of p-values close to one (Figure 4.3Ai) suggesting that the data has more variability than specified by the model. When sequencing batch was included as a factor in the design

matrix, the inflation of p-values close to one was reduced (Figure 4.3Aii), but fewer DEGs were identified using all tests (e.g. edgeR LRTr, FDR<0.1; group only, N=409; group + batch, N=142). This could suggest that some of the DEGs previously identified were false positives, but inspection of RLE distributions (Figure 4.1) and PCA (Figure 4.2A) shows that although samples vary by sequencing batch, there is additional variation within sequencing batches that may be due to other sources of unwanted variation (including RNA extraction and viral expression).

When two latent factors of unwanted variation (FUVs) identified by RUVg (Risso et al, 2014) were incorporated into the design matrix instead of batch, the number of DEGs detected was substantially increased (edgeR LRTr, FDR<0.1; N=794). Figure 4.3iii shows that more genes with low p-values were found and the distribution of genes with higher p-values was uniform. A similar distribution was obtained using DESeq2 with RUVg, as shown in Figure 4.3Aiv. After independent filtering by DESeq2, N=9216 genes were retained.

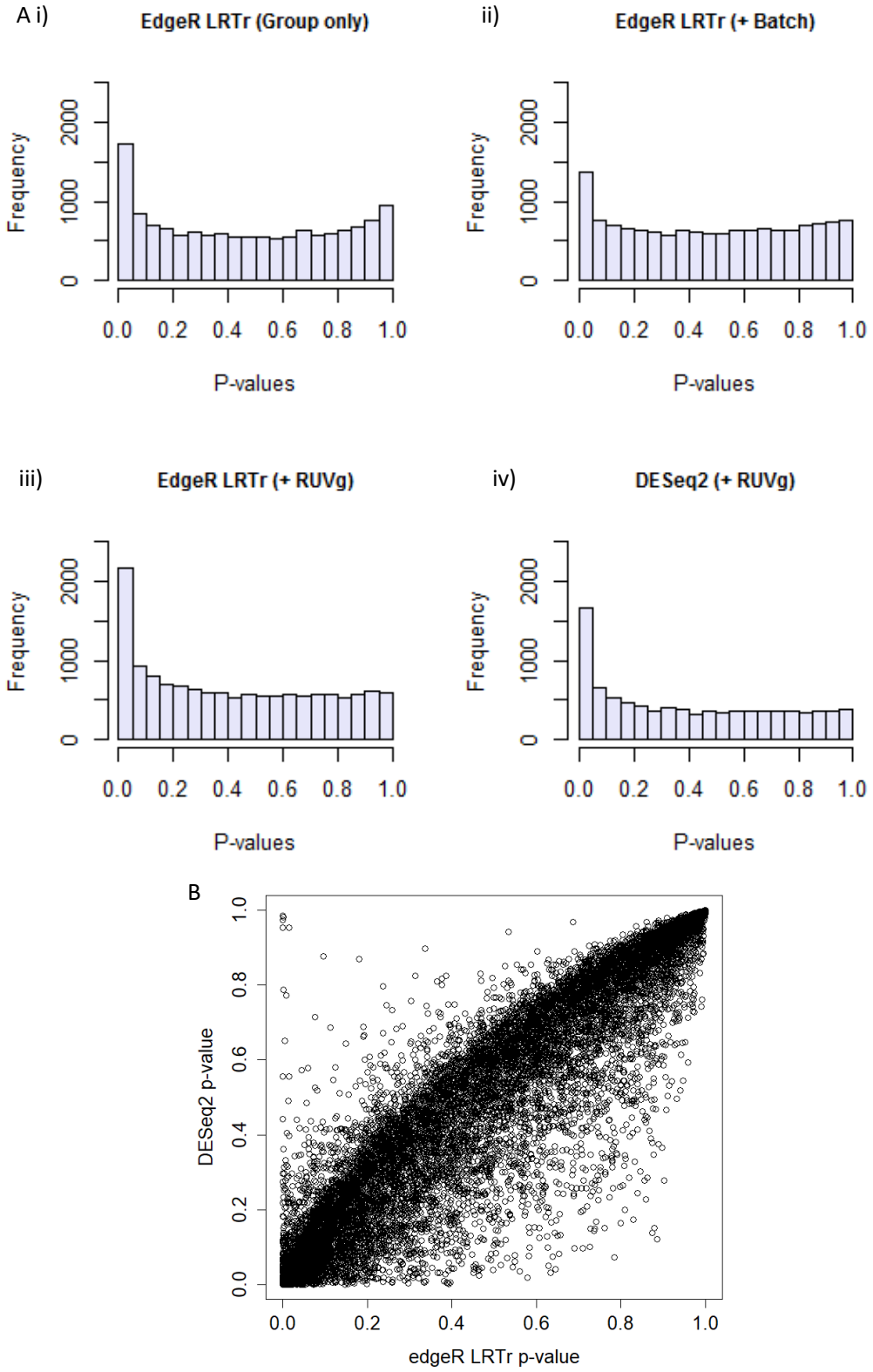


Figure 4.3: Comparison of differential expression analyses. All statistics refer to ANOVA-like tests for differential expression across all conditions. A: Histograms of P-values for edgeR robust likelihood ratio test (LRTr) with different design matrices and for DESeq2. i) EdgeR LRTr with experimental condition only. ii) LRTr with condition and batch. iii) LRTr with condition and two factors of unwanted variation (FUVs) from RUVg iv) DESeq2 with condition and two FUVs. B: Scatter plot showing the correlation between gene p-values (N = 13821) returned by the edgeR LRTr and DESeq2 pipelines, modelling condition and two FUVs in the experimental design. The Pearson correlation between tests was r = 0.923.

Table 4.2: Number of differentially expressed genes detected using different methods and thresholds

DE method		FDR (edgeR)/Adjusted P-value (DESeq2)	
		<0.05	***<0.1***
edgeR	***LRT***	544	910
	LRT (robust)	479	794
	QL	13	68
	QL (robust)	40	91
DEseq2	***LR (independent filtering)***	471	781

Table 4.2 shows the number of DEGs detected using five different methods of DE testing with correction for two factors of unwanted variation and FDR (edgeR)/adjusted p-value (DESeq2) thresholds of 5% and 10%. Relatively few genes were detected using the edgeR QL pipelines compared to the edgeR and DESeq2 LRT pipelines, and even fewer DEGs were identified using the QL F-test without adjustment for unwanted variation, suggesting that this method is not sufficiently sensitive for the present application. Robust dispersion estimation, to reduce the influence of outlier genes, increased the number of DEGs identified by the QL method and reduced the number of DEGs identified by the LRT method. Using the edgeR LRTr and DESeq2 methods, similar numbers of DEGs were identified, but just over 60% (N=482) of the genes identified using edgeR LRTr were also identified using DESeq2 at a threshold of FDR/adjusted p-value < 0.1. The identification of overlapping yet distinct subsets of DEGs by these tools is in accordance with previous findings (e.g. Soneson & DeLorenzi, 2013). The p-values returned by all pipelines were very highly correlated: the Pearson correlations between the four tests using edgeR were all greater than 0.965 and the correlation between the edgeR LRTr and DESeq2 tests was r=0.923 (see Figure 4.3B). Overall, N=54 DEGs overlapped between all five tests at a threshold of FDR<0.1 (see Table 4.7 for gene list) and these genes are analysed in the following section.

Using the edgeR LRTr pipeline, only one DEG (FDR<0.1) was identified by within-timepoint contrasts of novel and familiar conditions: TATA-box binding protein associated factor 13 (Taf13), a transcription factor, was more strongly expressed in the novel than the familiar condition at 10 minutes ($log_2FC = 1.62$, FDR = 0.00450).

4.3.1.3 Robust identification of differentially expressed genes

The DEGs detected by all five DE tests with an FDR (edgeR) or adjusted p-value threshold (DESeq2) of 0.1 (N=54; see section 4.3.1.2) are listed in Supplemental Table 4.7 with average log_2 CPMs and FDRs from edgeR LRTr analysis alongside gene symbols, Ensembl gene IDs (Aken et al, 2017), and full names. The expression profiles and functions of these DEGs are characterised below.

A heatmap of the DEGs identified in common between tests is presented in Figure 4.4. Most genes are either upregulated compared to other conditions at early timepoints and downregulated at late timepoints or downregulated at early timepoints and upregulated at late timepoints. The expression profiles of this set of DEGs are very similar between novel and familiar conditions at 90 minutes but diverge at 180 minutes: at 180 minutes, most genes appear to return to baseline expression in the novel condition whereas expression in the familiar condition is similar to expression in both conditions at 90 minutes. This pattern may reflect prolonged activity in the familiar condition or simply noise due to the low number of replicates at 180 minutes (N=2). At early timepoints, DEG expression is broadly similar in novel and familiar conditions and variation appears to be in the magnitude rather than the direction of changes among these genes.

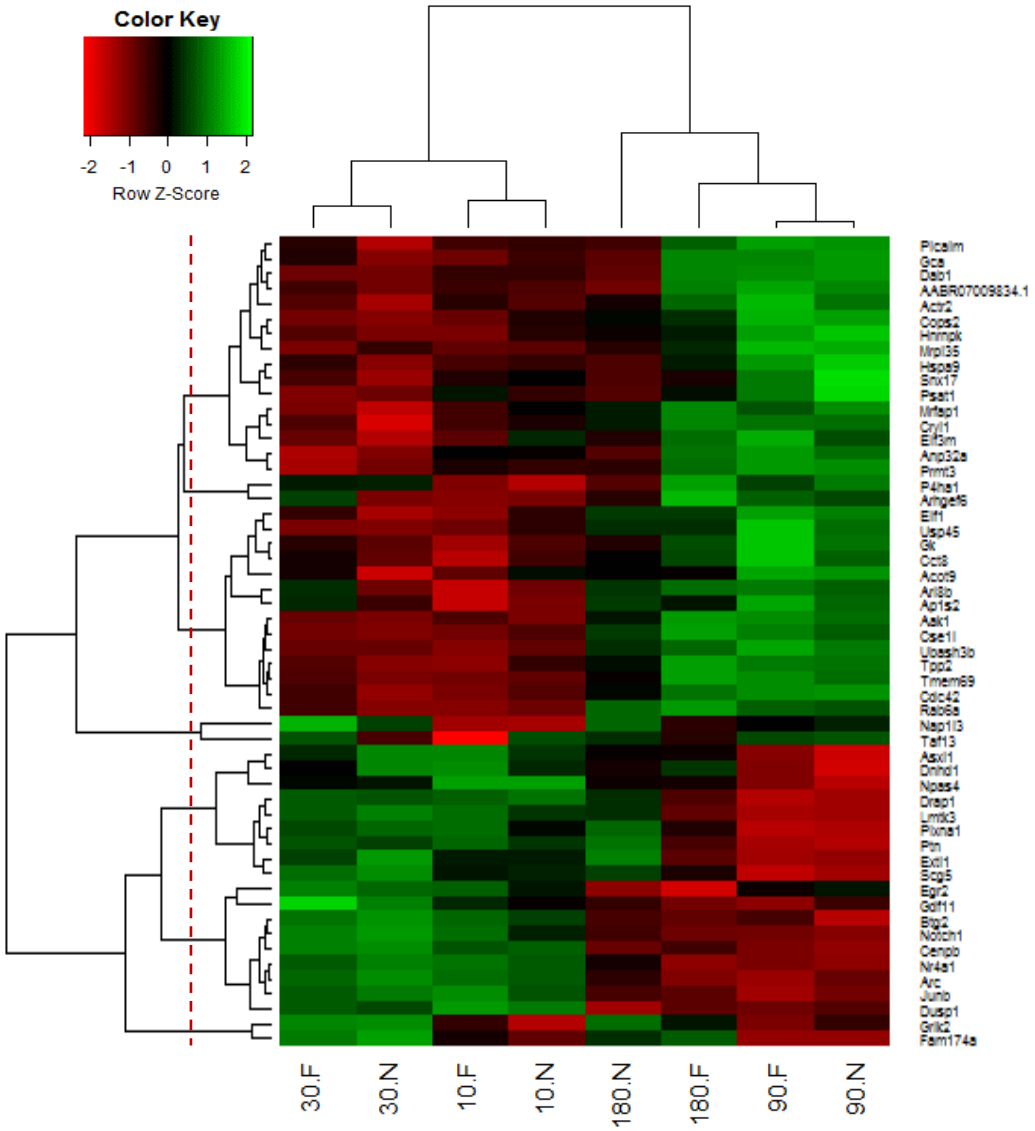


Figure 4.4: Heatmap of differentially expressed genes (DEGs) identified by all tests across all conditions with hierarchical clustering. The heatmap shows log_2 fold change Z-scores standardised for each gene (N=54). Expression in all conditions was compared to the no-image control condition using the edgeR robust likelihood ratio test. The complete-linkage method was used for hierarchical clustering of genes and conditions and the dissimilarity index was 1-correlation. Note that due to clustering, conditions are not

shown in order of time. A maximum distance between cluster elements of 0.5 is indicated by a dashed line on the gene dendrogram, showing that there are four clusters of >2 genes below this threshold with similar expression profiles. Novel and familiar conditions are most similar to the opposing condition at the same timepoint at 10, 30, and 90 minutes, whereas at 180 minutes, the familiar condition is most similar to the 90-minute conditions. Overall, DEGs have approximately opposing expression profiles at 10 and 30 minutes compared to 90 and 180 minutes.

4.3.1.4 Evidence for co-expression of functionally related genes

As genes with similar expression patterns are likely to be co-regulated (Eisen et al, 1998; Yeung et al, 2004; Iacono et al, 2013), average expression profiles of gene clusters can provide a more reliable and powerful indicator of differences between conditions than individual genes (Li & Bushel, 2016) and may compensate for some of the noise inherent in measurements of gene expression by RNAseq (Taub et al, 2010). Representative expression profiles of clusters of more than two genes with a maximum dissimilarity index of 0.5 between the least similar cluster elements were therefore compared to investigate expression differences between novel, familiar, and control conditions. The four clusters of genes that met these criteria are shown in Figure 4.5A (top) with representative expression profiles (bottom) showing standardised mean expression of each group of genes across conditions.

Gene expression profiles are relatively similar between novel and familiar conditions in all clusters and differ from the no-image control condition. There is also evidence of changes in expression over time in the control condition, as shown in previous studies (Cavallaro et al, 2002; Shires & Aggleton, 2008). In some cases, the same genes were found to have similar expression profiles in the control condition as in the novel and familiar conditions. This finding is unlikely to occur by chance as the control data was obtained from independent samples, was unreplicated, was not included in hierarchical clustering, and was pooled in the DE analyses by which the DEGs were identified. In addition, the relative expression of gene clusters is often similar at adjacent timepoints and the expression profiles are biologically plausible, despite the independence of samples from different timepoints. These findings suggest that the expression profiles are likely to represent meaningful trends in the data and may therefore be used to infer potential differences between conditions and over time.

In cluster 1 (which includes mainly IEGs, such as Arc and Junb), gene expression is higher at early timepoints relative to late timepoints. These genes are upregulated at 30 minutes in all conditions compared to other timepoints and expression at 10 minutes is higher in the novel and familiar conditions compared to the control and highest in the familiar condition. This indicates that early changes in gene expression are induced in all conditions, but that these changes may be faster and potentially greater in the experimental conditions than in the control. However, interpretation is complicated by the lack of a clear baseline. Expression of cluster 1 genes is similar in all three conditions at 90-180 minutes, which may correspond to basal expression, downregulation, or expression above basal levels in all conditions at all timepoints. Based on the known expression profiles of IEGs, which are upregulated upon neuronal stimulation and usually have low basal expression (Morgan et al, 1987; Lanahan & Worley, 1998), the observed expression profiles are most straightforwardly explained by expression close to basal levels at 90-180 minutes.

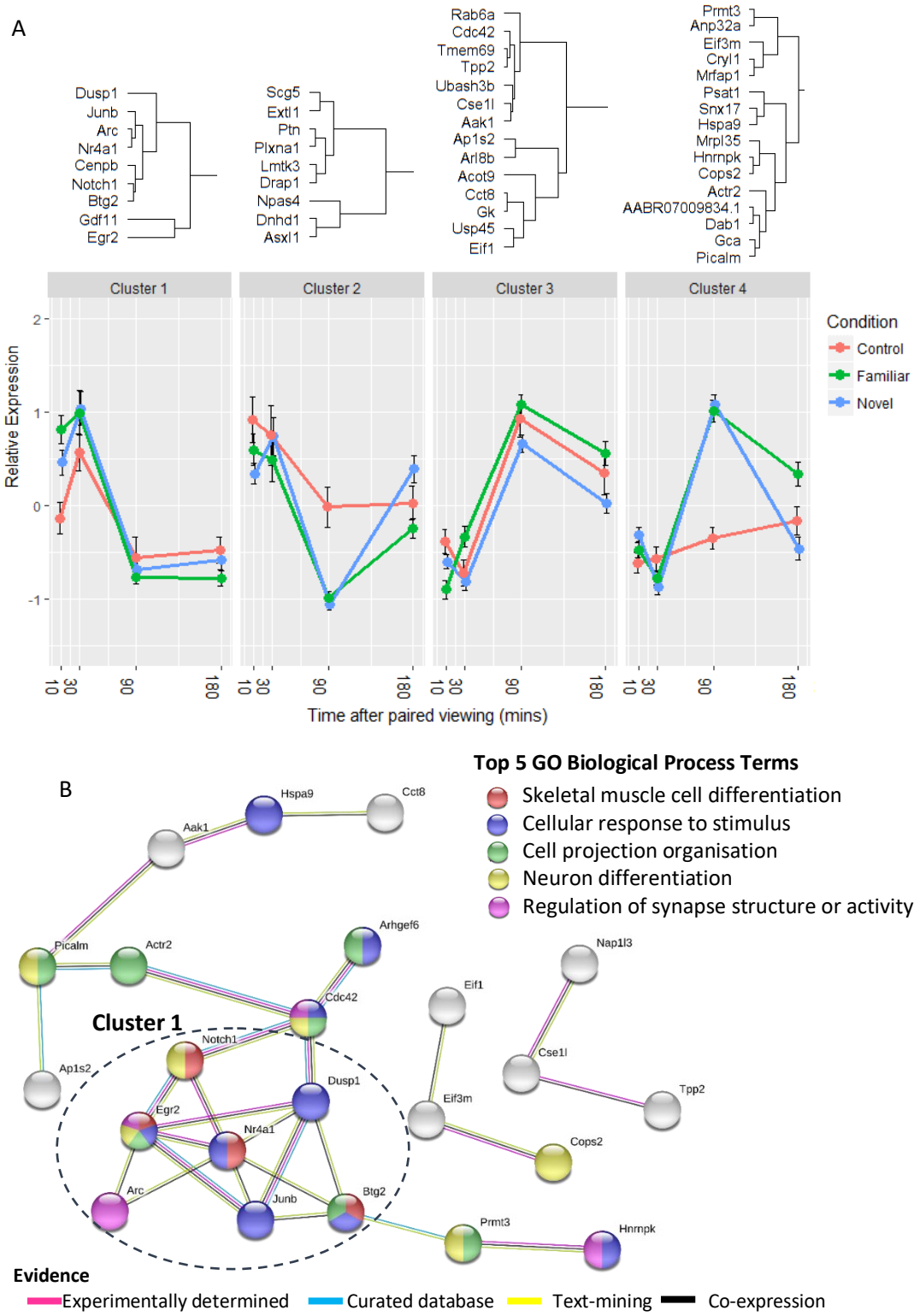


Figure 4.5: Representative expression profiles, protein-protein interactions, and functions of the differentially expressed genes (DEGs) identified by all tests (N=54). A: Top: Genes with similar expression profiles in hierarchical cluster analysis. Bottom: Cluster mean relative expression values across all conditions, calculated from per-gene z-scores of condition-mean CPMs. Error bars indicate standard error of cluster means. B: Protein-protein interactions between DEGs from the STRING database (von Mering et al, 2003) showing the sources of evidence. Node colour indicates genes annotated to the top five gene ontology (GO) biological process terms that were enriched among DEGs (N=53 in STRING) compared to a whole-genome background. Interactions were highly enriched among genes in cluster 1 (circled).

In addition to IEGs, cluster 1 contains Notch1, Gdf11, and Cenpb. Notch1 is a key transmembrane signalling protein that has been shown to be required for both LTP and LTD and is positively regulated by Arc (Alberi et al, 2011), while Gdf11 is a growth factor that has been linked to regulation of neurogenesis in the DG (Wu et al, 2003). Cenpb is a centromere protein localised to the nucleus (Earnshaw et al, 1987). It has been suggested that modification of centromere structure in response to neuronal activity and LTP induction (Billia et al, 1992; Wittmann et al, 2009) may contribute to the regulation of gene expression (Fujita & Yamashita, 2017).

The profiles of clusters 2 and 4 are distinguished by much lower or higher expression (respectively) at 90 minutes in the novel and familiar conditions compared to other conditions. Cluster 4 shows little change in the control condition over time and so this expression level is likely to correspond to basal expression, whereas the baseline in cluster 2 is less clear. Cluster 3 shows lower expression at early compared to late timepoints in all conditions, but a baseline for expression is again unclear and genes may have similar relative differences in expression caused by opposing patterns of regulation. A lower threshold for maximum within-cluster distance could be used to focus on genes with more similar expression profiles. However, this approach is impractical to extend to a larger dataset.

Among the 53 DEGs found in the STRING database (von Mering et al, 2003), 28 interactions were found, shown in Figure 4.5B, and protein-protein interactions were significantly enriched ($p = 3.8 \times 10^{-8}$). There is evidence of interactions between seven out of nine genes in cluster 1 in the STRING database (Figure 4.5B), indicating that these genes are strongly functionally related. Figure 4.5B also shows genes annotated to the top five enriched GO biological process terms among DEGs compared to a whole-genome background in STRING (listed in Table 4.3). The enriched terms are highly relevant to synaptic plasticity, including cellular response to stimulus (N=18; FDR=0.002), cell projection organisation (N=8; FDR=0.00313), and regulation of synapse structure or activity (N=6; p=0.00313) and each term represents a different subset of genes, with evidence of interactions between genes in all terms. Cluster 1 includes five genes annotated to cellular response to stimulus, consistent with the early upregulation of these genes, whereas cluster 4 contains 5 of the 11 genes annotated to cell projection organisation and/or neuron differentiation, functions consistent with the late upregulation of these genes.

Table 4.3: Top five enriched biological process GO terms among common differentially expressed genes

GO ID	GO Description	# Genes	FDR
GO:0035914	skeletal muscle cell differentiation	4	0.000358
GO:0051716	cellular response to stimulus	18	0.002
GO:0030030	cell projection organization	8	0.00313
GO:0030182	neuron differentiation	8	0.00313
GO:0050803	regulation of synapse structure or activity	6	0.00313

4.3.1.4.1 Correlated expression of post-synaptic density genes

The evidence for co-expression of functionally related genes in the dataset was further analysed by correlating the expression profiles of post-synaptic density (PSD) proteins across all samples (irrespective of differences between conditions). PSD genes were chosen for this analysis because they are important for signal transduction and plasticity at the synapse (Ziff, 1997) and have previously been shown to have highly correlated expression in neurons (Trinidad et al, 2013). From a list of known PSD proteins (N=110; Fernández et al, 2009; Trinidad et al, 2013), N=107 were expressed in CA1 TRAP samples, of which N=7 were differentially expressed across conditions (Glul, Ctnnb1, Grik2, Vdac3, Slc4a4, Grin1, Eef1a1; edgeR LRTr, FDR<0.1; circled in Figure 4.6).

Strong correlations were observed between the expression profiles of many of the PSD genes, as shown in Figure 4.6 (N = 55; $r > \pm 0.75$; median = 0.0279) and genes from the same family tended to be positively correlated, including Dlgap2-4, Actn1 & 2, Nlgn2 & 3, and Slc1a2 & 3. The most strongly correlated pair of PSD genes were the scaffolding proteins Shank1 and Shank3 (r = 0.934). Interestingly, two subsets of PSD genes were strongly negatively correlated: expression of Cltc, which codes for the clathrin heavy chain (the major component of intracellular vesicles), was negatively correlated with the glutamate receptor genes Grin1, Grin2b, and Grik5 (r = -0.849; -0.797; -0.881, respectively) and with Dlg4 (r = -0.851). Dlg4 codes for PSD protein 95 (Psd95), an adaptor protein that accounts for ~1% of protein content in the PSD (Sheng & Hoogenraad, 2007), and Dlg4 protein abundance has previously been shown to be highly correlated with the abundance of other PSD components in mouse forebrain PSD preparations (Trinidad et al, 2013). Here, expression of Dlg4 and glutamate receptor subunits is positively correlated with expression of genes coding for DLG-associated proteins (Dlg2-4) and membrane trafficking and actin-binding proteins, including Synpo, Syngap1, and Shank1 & 3. Genes that are negatively correlated with Dlg4 and positively correlated with Cltc are primarily associated with mitochondria and ATP metabolism, including Atp5b, Aco2, Glul (glutamine synthetase), and voltage-dependent anion channel (VDAC) isoforms 1-3, which are found in the mitochondrial outer membrane. These findings suggest that the synthesis of PSD plasticity proteins and PSD metabolism proteins may be co-ordinated.

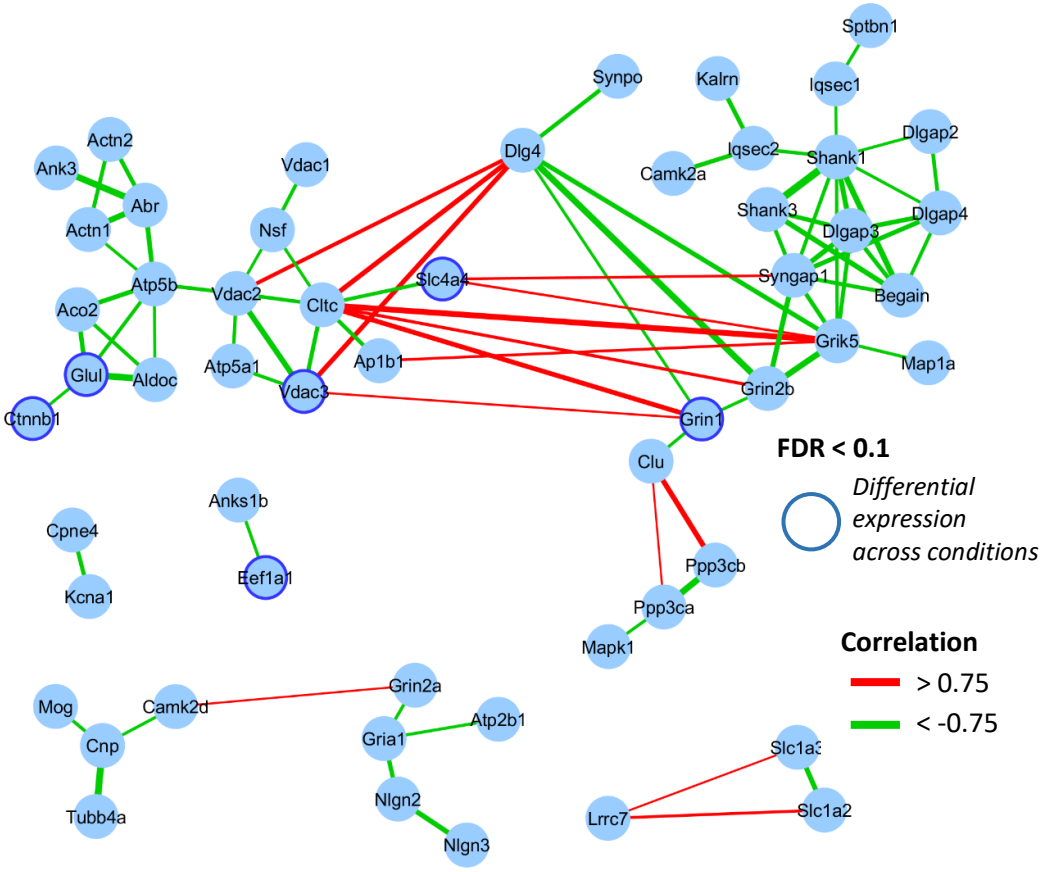


Figure 4.6: Correlated expression of postsynaptic density (PSD) proteins following paired viewing. Nodes represent PSD proteins identified in the present sample and edges represent strong correlations (> ±0.75) between expression profiles across all samples. Nodes with a dark border were found to be differentially expressed across conditions using the edgeR robust likelihood ratio test (false discovery rate < 0.1). Figure generated in Cytoscape v3.5.1.

4.3.2 Immediate early gene expression

IEGs are known to be expressed during memory formation (Guzowski et al, 2001) and are important for synaptic plasticity (Minatohara et al, 2016). IEG expression was therefore investigated as a positive control for changes in gene expression relevant to learning and memory processes.

The IEGs Arc, Fos, Egr1/Zif268, Junb and Npas4 are among the genes most strongly implicated in learning and plasticity in experimental work (Morgan et al, 1987; Abraham et al, 1991; Guzowski et al, 1999; Fleischmann et al, 2003; Strekalova et al, 2003; Ramamoorthi et al, 2011; Heroux et al, 2018). RUV-adjusted counts of these five IEGs are plotted in Figure 4.7. Except for Egr1, these IEGs were found to be significantly differentially expressed in response to paired viewing in the ANOVA-like test using the edgeR LRTr method described in section 4.3.1.2: Arc (FDR = 3.22×10^{-14}), Fos (FDR = 1.10×10^{-5}), Junb (FDR = 0.00048) and Npas4 (FDR = 3.22×10^{-14}). Figure 4.7 shows that all five IEGs have higher expression at early timepoints (10-30 minutes) compared to late timepoints (90-180 minutes) in the novel and familiar conditions. Arc, Fos, Junb and Egr1 are co-expressed in the novel and familiar conditions; Npas4 has a distinct expression

profile compared to the other IEGs and is sharply upregulated at 10 minutes compared to the other timepoints, but not at 30 minutes.

Some of the selected IEGs also show increases in expression at early timepoints in the no-image control condition: the expression of Npas4 peaks at 10 minutes, expression of Arc is increased at 10 and 30 minutes compared to 90 and 180 minutes and expression of Junb peaks at 30 minutes. Expression of Fos is relatively unchanged over time. Based on these preliminary data, expression changes in the no-image control condition are initiated later and/or are smaller in magnitude compared to the novel and familiar conditions. However, more replicates are needed to confirm these findings.

Overall, Figure 4.7 shows that the expression of key IEGs changes over time in both the novel and familiar conditions; some IEGs show similar directional changes in expression of lower magnitude in the control condition.

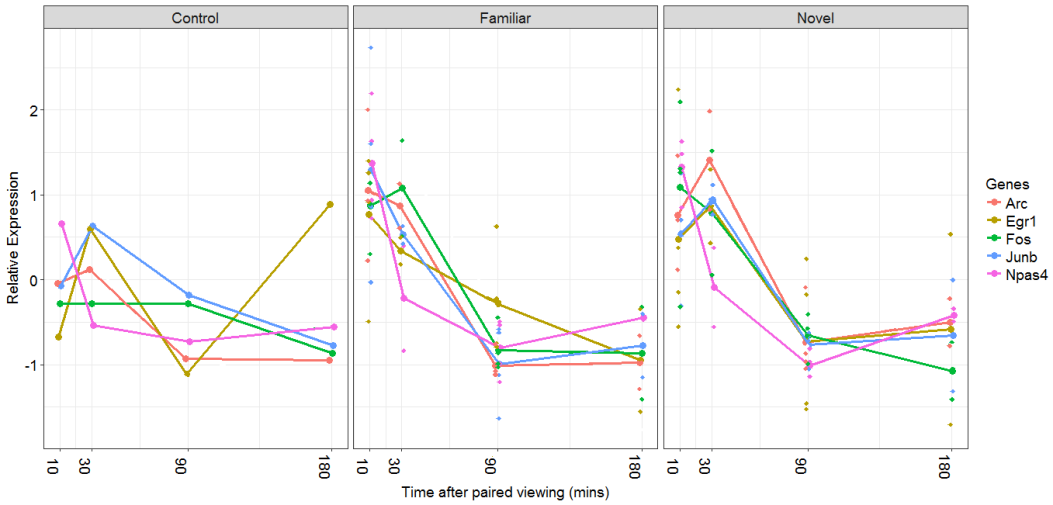


Figure 4.7: Expression of key immediate-early genes (IEGs) in PV-TRAP experiment. Mean (large points connected by lines) and individual sample (smaller points) RUV-adjusted counts for Arc, Egr1 (Zif268), Fos, Junb, and Npas4 in novel, familiar, and control conditions 10, 30, 90, and 180 minutes after test sessions.

A list of IEGs and delayed primary response genes (i.e. genes that do not require de novo protein synthesis for their expression but are induced later than IEGs; Tullai et al, 2007) was obtained from Saha et al (2011). A heatmap of primary response gene expression in the novel and familiar conditions at each timepoint compared to no-image controls is shown in Figure 4.8, excluding those genes with $p < 0.1$ in the ANOVA-like test to reduce noise. Ten of the selected genes were found to be differentially expressed in response to paired viewing (Arc, Npas4, Fos, Nr4a1, Dusp1, Egr2, Junb, Btg2, Ppp1r15a, and Bdnf). Hierarchical clustering of all conditions and timepoints shows that the greatest difference in primary response gene expression is between early and late timepoints (as also shown by PCA; Figure 4.2): primary response gene expression is more similar in novel and familiar conditions at the same timepoint than at adjacent timepoints in the same condition. As expected, most of the genes have higher expression at early timepoints. Novel and familiar conditions are more similar at the 10 and 90 minute timepoints than at the 30 and 180 minute timepoints, likely due to the greater number of replicates at 10 and 90 minutes. Some of the primary response genes have very

similar expression profiles (in particular, Nr4a1, Arc, Ppp1r15a and Junb), which suggests they are co-regulated. A subset of genes (Mbnl2, Nup98 and Arf4) shows a trend towards upregulation at later timepoints, with peak expression at 90 minutes. Interestingly, three genes (Dusp5, Gadd45b and Rasl11a) show a trend towards greater expression changes in the familiar condition than in the novel condition. Another subset of genes (Egr2, Trib1 and Kcnf1) shows a trend towards earlier upregulation in the familiar condition.

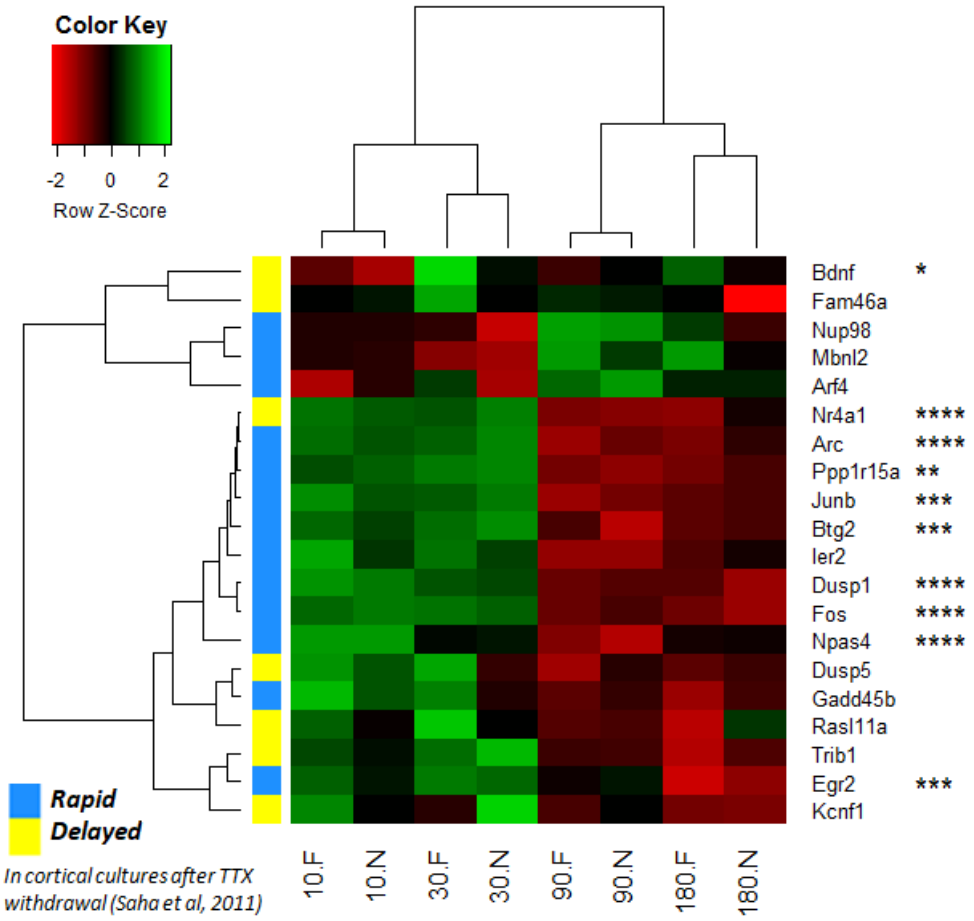


*Figure 4.8: Differential expression of immediate early genes (IEGs) and delayed primary response genes. Heatmap shows log_2 fold change Z-scores standardised for each gene. Expression in all conditions was compared to the no-image control condition using the edgeR robust likelihood ratio test. Genes with p<0.1 were excluded from the heatmap to reduce noise and differentially expressed genes (N = 10; FDR < 0.1) are indicated by asterisks: * FDR <0.1; ** <0.01; *** <0.001; **** <0.0001. The complete-linkage method was used for hierarchical clustering of genes and conditions and the dissimilarity index was 1-correlation. The gene list was obtained from Saha et al (2011) and classification of each gene as rapid or delayed in that study are indicated by the coloured bar to the left of the heatmap.*

In Figure 4.8, hierarchical clustering of primary response genes is compared with the classification of genes as rapid (IEGs) or delayed in a previous study, based on upregulated expression in cortical cultures after tetrodotoxin withdrawal at 15 minutes, which was dependent on stalling of RNA Pol II, or 45 minutes respectively (Saha et al, 2011). All but one (Nr4a1) of the nine genes with upregulated expression in both novel and familiar conditions at early timepoints in the present study was previously found to be rapidly expressed, as were the three genes here strongly upregulated at 90 minutes

(Saha et al, 2011). Interestingly, most of the genes with a similar response in both novel and familiar conditions were rapidly expressed in cortical cultures and regulated by promoter proximal Pol II stalling, whereas most of the genes here trending towards different expression profiles in novel and familiar conditions were found to have delayed expression by Saha et al, suggesting that underlying regulatory mechanisms may account for these differences.

Overall, these results show an increase in the expression of genes previously reported to be involved in learning and memory formation in the thirty minutes following the presentation of image-arrangements compared to later timepoints, and a reduced response in the corresponding no-image control condition; a small number of known primary response genes showed trends towards different dynamics in the novel and familiar conditions.

4.3.3 Exploratory analyses of genes that are differentially expressed in response to the paired viewing protocol

This section presents exploratory analyses of the DEGs identified by the edgeR LRTr ANOVA-like test for differential expression between all conditions (N=794; FDR<0.1), including gene ontology, known protein-protein interactions, and k-means clustering.

4.3.3.1 Differentially expressed genes

To provide an overview of the functions of DEGs, PANTHER protein classes are presented in Figure 4.9A. The most strongly represented categories, each including more than 50 genes, are enzyme modulators, hydrolases, and nucleic acid binding proteins. Combined with strong expression of transferases and transcription factors, these results suggest that the regulation of gene expression and enzymatic activity are particularly important functions modulated between these conditions. Additionally, several categories of proteins involved in synapse remodelling are expressed, including cytoskeletal proteins, membrane traffic proteins, receptors, and cell adhesion molecules.

A heatmap of DEGs is presented in Figure 4.9B for an overview of differential gene expression across conditions. As shown in the previous section, expression is mainly regulated in the same direction between paired samples, except at 180 minutes. Overall, expression of DEGs is broadly similar at 10 and 30 minutes and reversed at 90 minutes. Overall, more DEGs are upregulated at early timepoints than at later timepoints. Hierarchical clustering of the fold changes across conditions compared to controls does not divide genes into distinct clusters using the complete-linkage method.

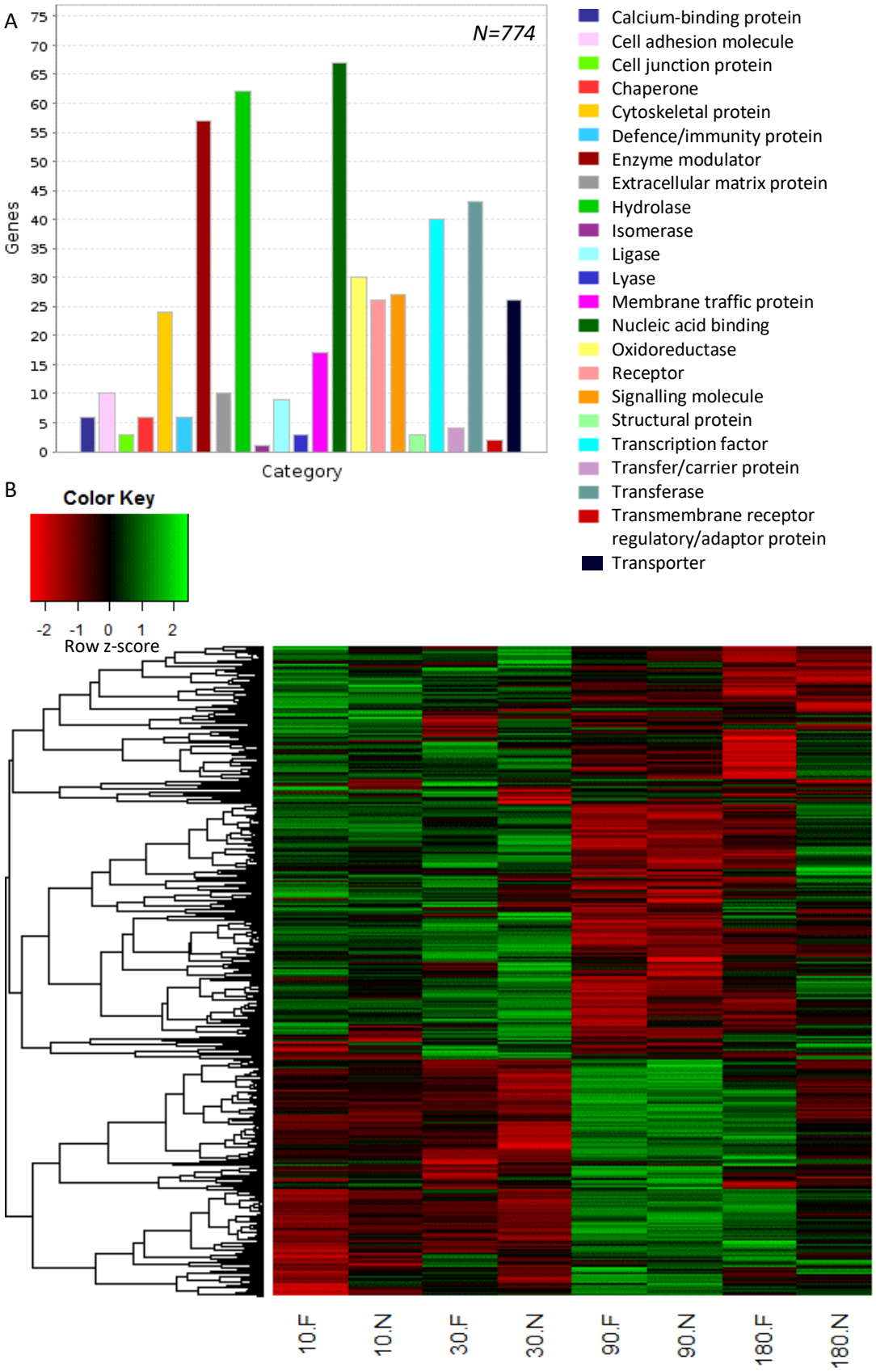


Figure 4.9: Differentially expressed genes (DEGs) identified by edgeR robust likelihood ratio test across all conditions (N=794; FDR<0.1). A: Number of DEGs annotated to PANTHER protein classes. N=774 genes were found in the PANTHER database. B: Heatmap of DEGs with hierarchical clustering of gene expression profiles. The heatmap shows log_2 fold change Z-scores standardised for each gene. Expression in all conditions was

compared to the no-image control condition. The complete-linkage method was used for hierarchical clustering of genes and the dissimilarity index was 1-correlation. Conditions are presented in order of time.

4.3.3.2 Protein-protein interactions between differentially expressed genes

Evidence for protein-protein interactions between genes that were differentially expressed in response to paired viewing was found using the STRING database, which showed that interactions were strongly enriched among DEGs (nodes = 722; edges = 1374; $p < 1.0x10^{-16}$).

Strong interactions (combined interaction score > 0.7) are shown in Figure 4.10 and at this threshold, a large network with two densely connected subnetworks was formed connecting 212 DEGs. Genes within the larger subnetwork include IEGs (e.g. Fos, Bdnf, Egr2) as well as genes involved in signal transduction (Cdc42, Rab6a), actin remodelling (Actr2, Sept2), and vesicle endocytosis (Picalm, Aak1, Synaptophysin (Syp)). Genes in the smaller subnetwork are primarily involved in the regulation of gene expression. In addition to ATP synthesis by mitochondria (e.g. Cox2, Atp5s), functions include transcription and DNA repair (Taf13, Rfc4), mRNA splicing (Hnrnpk, Snrnp70), translation (Eif1, Eef1a1, Rps8), and protein degradation (Psma3, Ube2a).

Genes with betweenness centrality scores (C_b; Brandes, 2001) within the highest 5% of the main network (N = 11; $0.107 < C_b < 0.435$) were identified as putative hub genes (see Figure 4.11A and Table 4.4). The distribution of betweenness centrality indices by number of neighbours (Figure 4.11B) shows that for most genes in the main network, betweenness centrality increases approximately linearly with node degree. However, four nodes (Ccnh, Phka1, Ppp1cb, Rock2) have much higher betweenness centrality scores than predicted by this trend as these nodes provide the shortest route connecting the two subnetworks.

All putative hub proteins within the larger subnetwork are known to be key mediators of synaptic plasticity: the NMDAR subunit GluN1 (Grin1), the calmodulin Calm2, β-catenin (Ctnnb1), the Rho GTPases Rac1 and Cdc42, and the Rho-kinase Rock2. GluN1 is the essential subunit of NMDARs and its expression in CA1 is also required for the consolidation of long-term spatial memory (Shimizu et al, 2000). During the induction of synaptic plasticity, calmodulin (CaM) binds to Ca^{2+} to activate Ca^{2+}/CaM-dependent kinases and phosphatases, including CaMKii and protein phosphatase 3 (calcineurin). CaM is therefore an essential signal integrator for the conversion of synaptic activity into downstream cellular processes (Xia & Storm, 2005). β-catenin has similarly been described as a 'hub' for neuronal plasticity (Maguschak & Ressler, 2012) due to its roles in both cell-cell adhesion at synapses (Takeichi & Abe, 2005) and signalling to the nucleus to initiate transcription as part of the Wnt signalling pathway (Chen et al, 2006; Maguschak & Ressler, 2012). Expression of β-catenin in the amygdala has been shown to be required for consolidation of fear memory (Maguschak & Ressler, 2008). Rho GTPases are signalling proteins that regulate actin (Hall, 1998) and previous studies have found different effects of Rac1, Cdc42, and Rock2 on structural plasticity. Cdc42 knockdown has been found to impair the formation of spines and synapses in hippocampal neurons (Wegner et al, 2008), whereas knockout of Rac1 has been found to increase spine size while reducing Psd95 density (Haditsch et al, 2009). In Rock2

knockout mice, a reduced number of excitatory synapses, altered spine morphology, and impaired basal transmission have been reported (Zhou et al, 2009).

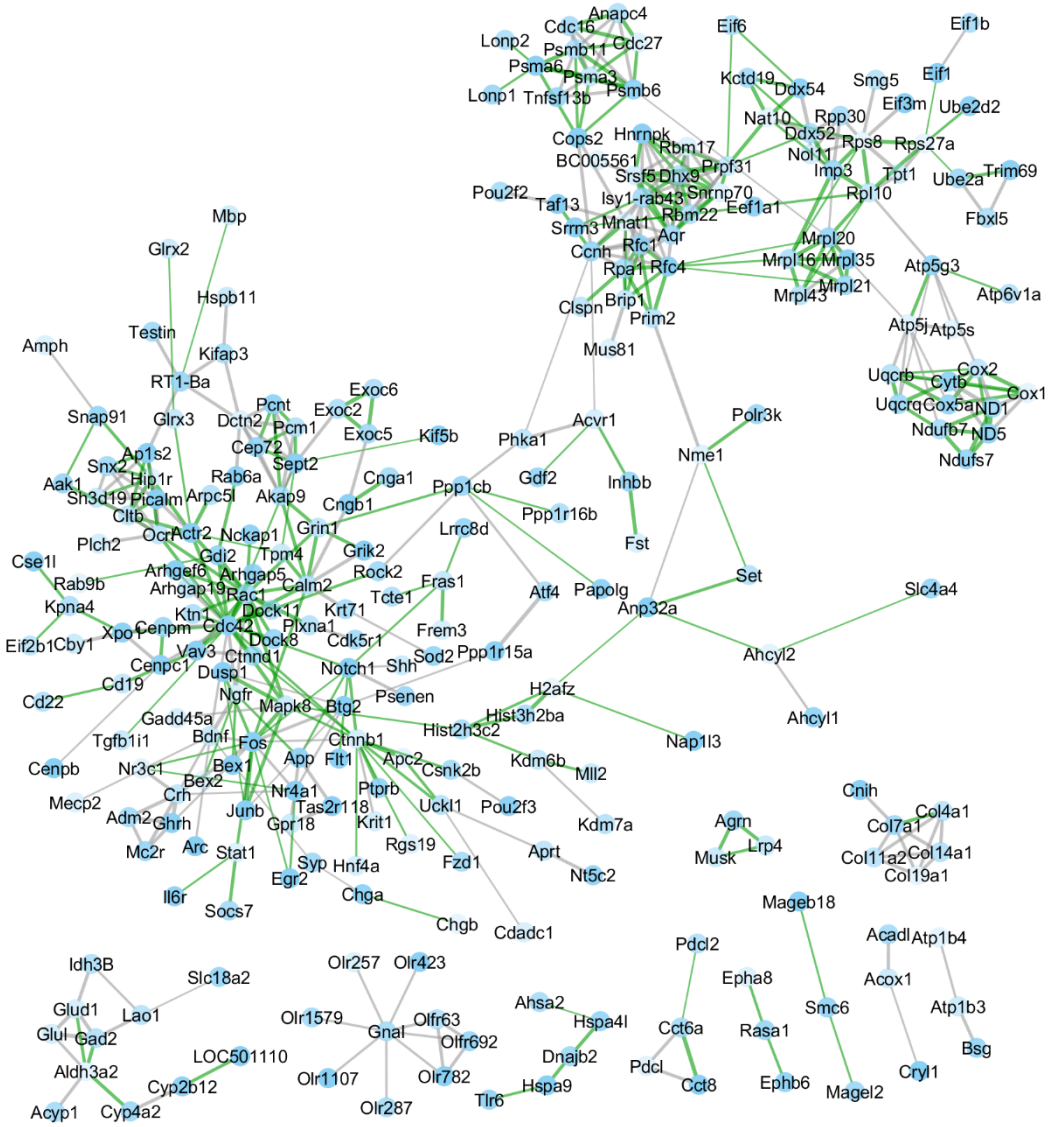


Figure 4.10: Network of differentially expressed genes (DEGs) following paired viewing. Nodes are genes found to be differentially expressed across conditions (false discovery rate (FDR) < 0.1) that have strong evidence of interactions with other DEGs and edges represent STRING combined interaction scores > 0.7. Node transparency is mapped to FDR with opaque nodes having lower FDRs. Thicker edges indicate combined scores closer to one and edges coloured green have experimental interaction scores > 0. Network generated in Cytoscape 3.5.1, applying the Prefuse force-directed layout.

Predicted hub proteins within the smaller subnetwork are cyclin H (Ccnh), replication factor C subunit 4 (Rfc4), and the mitochondrial ribosomal protein Mrpl20. Cyclins control cell-cycle progression by the activation of cyclin-dependent kinases (Cdks) and proteins of both families have been found to play alternative roles in synaptic plasticity in post-mitotic neurons (Frank & Tsai, 2009). Rfc4 is part of the DNA replication factor C clamp-loading complex and interacts with other DEGs required for DNA replication and repair including Prim2, Brip1, and Rpa1. These proteins may be synthesised to repair double-strand breaks in DNA; double-strand breaks have recently been found to facilitate rapid transcription of IEGs in response to neuronal activity and are repaired within two hours (Madhabushi et al, 2015). Synthesis of Mrpl20 and other mitochondrial ribosome proteins may be required to meet the increased energy demands during plasticity by increasing mitochondrial gene expression. In a recent study, hemizygous deletion of Mrpl40 was found to impair short-term potentiation and working memory by impairing the calcium buffering function of mitochondria (Devaraju et al, 2017).

The two subnetworks in Figure 4.11 are connected by protein phosphatase 1 (PP1) catalytic subunit β (Ppp1cb), which has the highest betweenness centrality score in the network (C_b = 0.434), and phosphorylase kinase subunit alpha (Phka1). PP1 and other protein phosphatases are key regulators of plasticity and interact with many different proteins, dependent on subunit composition (Greengard et al, 1999). Expression of PP1 sustains LTD and favours forgetting (Morishita et al, 2001), whereas inhibition of PP1 is required for LTP (Blitzer et al, 1998). Inhibitors of PP1 are also differentially expressed, such as Ppp1r16b. Phka1 is a Ca^{2+}/CaM-dependent kinase also regulated by cAMP (Lukas et al, 1998). Although there has been relatively little research into the contribution of Phka1 to learning and memory, phosphorylation of Phka1 in dCA1 has been found to be significantly reduced for the first hour following LTD-inducing stimulation (Thiels et al, 1998). The subnetworks are also connected by a second route via Anp32a, an inhibitor of protein phosphatase 2A (PP2A; Figure 4.10). LTD has been found to transiently activate PP1 while persistently activating PP2A (Thiels et al, 1998).

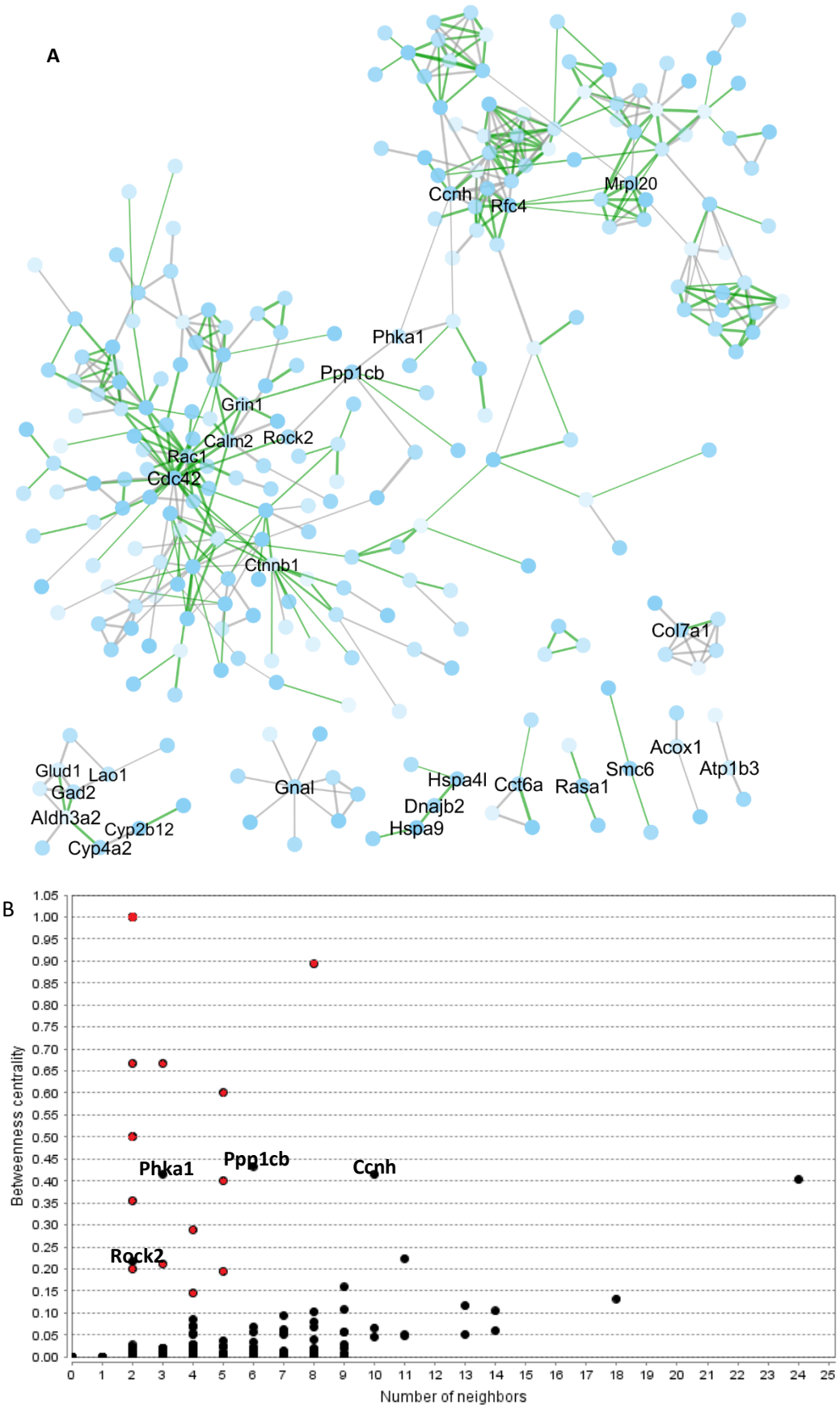


Figure 4.11: Putative hub genes within the network of differentially expressed genes (DEGs) following paired viewing. A: Nodes represent genes found to be differentially expressed across conditions (false discovery rate (FDR) < 0.1) that have strong evidence of interactions with other DEGs and edges represent STRING

combined interaction scores > 0.7. Node transparency is mapped to FDR with opaque nodes having lower FDRs. Thicker edges indicate combined scores closer to one and edges coloured green have experimental interaction scores > 0. The Prefuse force-directed layout was applied. B: Distribution of betweenness centrality scores by number of neighbours. Nodes not connected to the main network are indicated in red. Within the main network, four genes (Ccnh, Phka1, Ppp1cb, Rock2) have higher betweenness centrality scores than expected based on node degree. Graphs generated in Cytoscape 3.5.1.

Smaller groups of at least three proteins with strong predicted interactions are also shown in Figure 4.11 and the central node in each group of five or more proteins is listed in Table 4.4. The largest of these groups contains genes involved in the metabolism of glutamate and fatty acids, including cytochrome P450 4a2 (Cyp4a2) and aldehyde dehydrogenase Aldh2a3. Lipid metabolism has been shown to modulate synaptic signalling and plasticity (Koudinov & Koudinova, 2001; Bazan, 2003) and impaired lipid metabolism is associated with deficits in spatial learning and LTP at CA3-CA1 synapses (Stranahan et al, 2008) and implicated in Alzheimer's disease pathology (Hooijmans & Kiliaan, 2008). Knockout of Aldh2 has been found to have different effects on MWM performance dependent on age (Ohsawa et al, 2008; Jamal et al, 2012). Another group of DEGs are olfactory receptors, a family of G-protein coupled receptors that activate G-protein alpha (encoded by Gnal) and are known to have functions unrelated to olfaction in neurons, such as Ca^{2+} signalling and homeostasis (Ferrer et al, 2016). Mutations in Gnal have been implicated in the pathology of bipolar disorder and attention deficit hyperactivity disorder (Berrettini et al, 1998; Laurin et al, 2008). Two additional groups of DEGs are collagens and members of the heat shock protein 70 (Hsp70) family. Collagens are a large family of extracellular matrix proteins and one of the collagen genes here found to be differentially expressed, Col19a1, has been shown to contribute to the formation of hippocampal synapses (Su et al, 2010). Hsp70 expression correlates with metabolic activity and has previously been shown to increase in response to spatial learning (Pizarro et al, 2003).

Table 4.4: Putative hub genes in the network of differentially expressed genes following paired viewing

Gene	Name	FDR	Degree	Betweenness Centrality
Aldh3a2	Fatty aldehyde dehydrogenase	0.097	5	0.600
Calm2	Calmodulin 2	0.048	9	0.107
Ccnh	Cyclin-H	0.022	10	0.414
Cdc42	Cell division control protein 42 homolog	0.004	24	0.403
Col7a1	Collagen type VII alpha 1 chain	0.030	5	0.400
Ctnnb1	Catenin beta-1	0.080	13	0.116
Dnajb2	DnaJ heat shock protein family (Hsp40) member B2	0.030	2	0.667
Gnal	Guanine nucleotide-binding protein G(olf) subunit alpha	0.037	8	0.893
Grin1	Glutamate receptor ionotropic, NMDA 1	0.059	5	0.195
Mrpl20	Mitochondrial ribosomal protein L20	0.024	9	0.159
Phka1	Phosphorylase b kinase regulatory subunit alpha	0.058	3	0.416
Ppp1cb	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	0.022	6	0.434
Rac1	Ras-related C3 botulinum toxin substrate 1	0.037	18	0.132
Rfc4	Replication factor C (Activator 1) 4	2.30×10^{-4}	11	0.224
Rock2	Rho-associated protein kinase 2	0.031	2	0.217

4.3.3.3 K-means clustering

Gene relative expression profiles (as in Figure 4.5A, using RUV-adjusted data) were clustered using the k-means clustering algorithm (Hartigan & Wong, 1979) to explore underlying patterns in the dataset, such as the timing of expression changes and the relationship between expression timing and function, by generating a reduced represention of the dataset.

The optimum number of clusters (k) in the novel and familiar conditions was determined by the first maximum of the gap statistic (Figure 4.12A) and was found to be k=3 for the familiar data and k=4 for the novel data. When the novel and familiar datasets were combined, the optimum number of clusters was k=8, suggesting that although gene expression in paired conditions is similar compared to conditions from other timepoints, the information in paired samples is not redundant and overall gene expression profiles differ between conditions. Subsequent clustering was therefore performed separately on novel and familiar datasets.

The results of k-means clustering are presented in Figure 4.12B, which shows the mean relative expression profiles of each cluster of genes in the novel and familiar conditions over time. The total within-SSE was 549 for the novel clusters (range 105-159) and 702.9 for the familiar clusters (range 213-266). Approximately two thirds of the overall variance in the data was explained by the clustering configuration (Novel: between-SSE/total-SSE = 0.661; Familiar: between-SSE/total-SSE = 0.621). Figure 4.12B shows that there are some clusters with similar expression profiles in the two conditions: clusters F1 and N3 are expressed strongly at early timepoints, particularly at 30 minutes, and clusters F3 and N1 are most strongly expressed at 10 minutes and show a gradual reduction over time. The latter profile could be plausibly be accounted for by upregulation at 10 minutes or repression from 30-180 minutes. In the familiar condition, the remaining cluster of genes (F2) has increasing expression over time from 10 to 90 minutes and sustained expression at 180 minutes. In the novel condition, clusters N2 and N4 have fluctuating and roughly opposing expression profiles: expression of N2 is lowest at 90 minutes and peaks at 180 minutes, whereas N4 is most highly expressed at 90 minutes and returns to (assumed) baseline at 180 minutes.

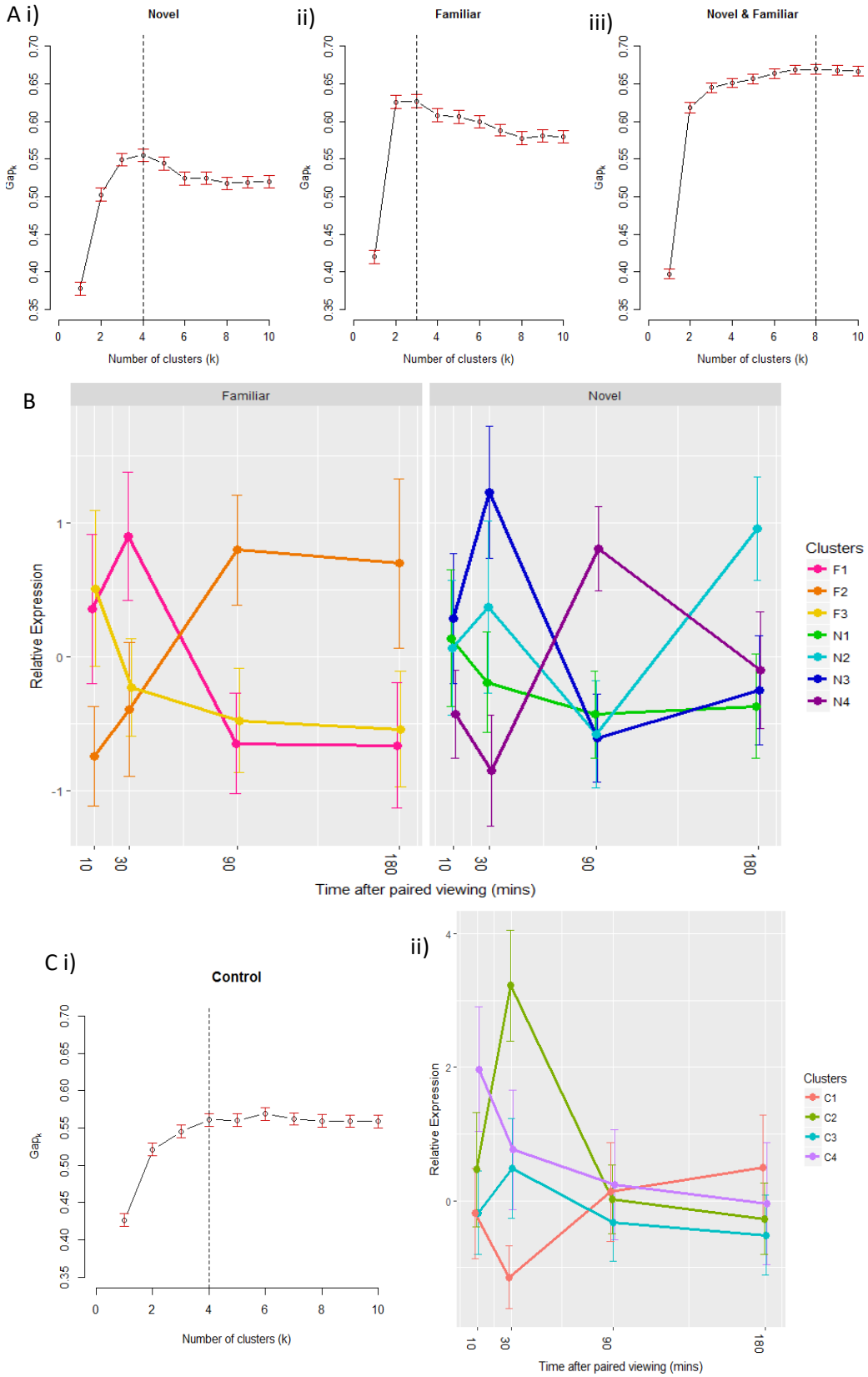


Figure 4.12: K-means clustering of expression profiles of differentially expressed genes by stimulus condition. K-means clustering was performed on standardised mean RUV-adjusted counts for each condition. A: The optimal number of clusters in novel (i), familiar (ii), and novel and familiar combined (iii) datasets was determined by plotting the gap statistic for values of k between 1 and 10. The gap statistic compares the total within-cluster sum-of-squared errors (SSE) to the mean total within-SSE of bootstrapped uniformly-distributed reference samples (N=500). The optimal number of clusters was defined as the first maximum

(marked by dashed line). Error bars indicate standard error (SE) of total within-SSE for bootstrapped samples. B: Mean and standard deviation of relative expression values at each timepoint for genes in clusters found in novel and familiar datasets. C: Optimisation (i) and expression profiles (ii) of clusters in control data as described above for novel and familiar data.

To show the gene set overlap between novel and familiar clusters, a contingency table is presented in Table 4.5A. There is almost complete overlap between F2 and N4, most likely due to the upregulation of a similar set of genes in both conditions at 90 minutes. Clusters F1 and F3 have the greatest overlap with clusters N3 and N1 respectively and these clusters also have the most similar temporal expression profiles. However, both pairs of clusters also contain many distinct genes that do not overlap between clusters. The smallest cluster, N2, overlaps with both F1 and F3.

K-means clustering was also performed on the no-image control data to determine whether similar expression profiles would be found. The optimum number of clusters was determined to be k=4, as shown in Figure 4.12C(i). The total within-SSE after clustering was 1600.9 (range 331-500). Control data was much more variable overall (the total-SSE was 4241, compared to 1619 and 1854 in novel and familiar data respectively), which may be due to the lack of replicate samples, but the clustering configuration nevertheless accounted for a similar proportion of overall variance (62.3%).

The mean relative expression profiles of clusters in the no-image control data are presented in Figure 4.12C(ii). In all four clusters, mean relative expression is roughly constant and close to zero at 90 and 180 minutes, suggesting that changes in expression do not occur at these timepoints in the control condition and implying that mean relative expression values close to zero are likely to be a useful proxy for baseline expression here. Gene expression changes are observed at early timepoints: C2 expression peaks strongly at 30 minutes, C4 peaks at 10 minutes and slowly decays to baseline, and C1 has lower expression at 30 minutes. Mean relative expression of C2 at 30 minutes is much higher than expression in any other condition or cluster, which may be due to noise caused by amplification and a lack of replication.

Table 4.5: Contingency tables for genes with clustered expression profiles in novel, familiar, and control data, showing number of genes in each cluster and gene set overlap of clusters in different conditions

A		***Novel clusters***				**Total # Genes**
		1	*2*	*3*	*4*	
Familiar clusters	*1*	73	60	**111**	6	*250*
	2	8	7	3	**261**	*279*
	3	**120**	42	90	13	*265*
Total # Genes		*201*	*109*	*204*	*280*	***794***

B		***Control clusters***				**Total # Genes**
		1	*2*	*3*	*4*	
Novel clusters	*1*	17	**94**	45	45	*201*
	2	17	20	44	28	*109*
	3	7	51	**112**	34	*204*
	4	**232**	3	28	17	*280*
Total # Genes		*273*	*168*	*229*	*124*	***794***

C		***Control clusters***				**Total # Genes**
		1	*2*	*3*	*4*	
Familiar clusters	*1*	20	66	**115**	49	*250*
	2	**230**	4	29	16	*279*
	3	23	**98**	85	59	*265*
Total # Genes		*273*	*168*	*229*	*124*	***794***

The gene set overlap between clusters in the control condition and clusters in the novel and familiar conditions was investigated using contingency tables, presented in Table 4.5. Cluster C1, which shows a trough in expression at 30 minutes, has very strong overlap with clusters N4 and F2, which had peak expression at 90 minutes in their respective conditions. The clustering of these genes in the control condition may be related to the inhibition of plasticity processes or may be simply an artefact caused by the standardisation of gene expression values across all conditions. The remaining three clusters in the control data did not straightforwardly map onto clusters in the novel and familiar data. The cluster of genes that were strongly expressed at 30 minutes in the control data, C2, has the greatest overlap with clusters in the novel and familiar conditions with peak expression at 10 minutes (N1 and F3). C3, which shows little change in expression over time, was enriched for genes in N3 and F1, which are upregulated at early timepoints. Overall, although some of the clusters in the control condition have similar expression profiles to clusters in the novel and familiar conditions, these clusters represent different sets of genes.

4.3.3.4 GO enrichment mapping

To determine whether clusters of genes with different expression profiles were involved in related functions, GO analyses were conducted on each cluster of genes found in the novel and familiar conditions. A network of enriched GO terms was produced to visualise enriched functions and functional relationships within and between clusters of genes with similar expression across conditions (Figure 4.13). For simplicity, in this section cluster expression profiles will be referred to by the timepoint of peak expression. A detailed map of the functions of groupings of enriched GO terms is presented in Figure 4.13A, summarising GO terms that describe overlapping sets of genes and related cellular processes. A summary map is presented in Figure 4.13B, showing three broad groupings of GO terms based on network topology, functional relationships, and the subset of clusters in which those GO terms are enriched.

Figure 4.13 shows that the two most strongly related clusters of genes in the novel and familiar data were F2 (peak 90-180 minutes) and N4 (peak 90 minutes), as identified in the previous section by gene set overlap (Table 4.5). Enriched GO terms in these two clusters are highly distinct from those enriched in other clusters within the GO map and include translation and ribosomal constituents, ATP synthesis, other metabolic processes, enzyme activity, and terms related to synaptic remodelling including microtubule activity, actin cytoskeleton, leading edge and spine formation. Among these functions, terms related to metabolism and translation are more strongly enriched in the novel condition, whereas ATP synthesis terms are more strongly enriched in the familiar condition.

GO terms enriched in the two clusters of genes with highest expression at 10 minutes (F3 and N1) are mainly related to activity at the synapse and the cellular response to stimuli, suggesting that these genes are likely to be upregulated at 10 minutes. Although similar functions are enriched in these clusters, most of the enriched GO terms are different. For example, terms related to growth are enriched in both clusters, but in the novel condition the enriched terms are specifically related to positive regulation of cell growth, whereas the terms enriched in the familiar condition are more general. GO terms related to the cell surface are also specifically enriched in N1. In contrast, F3 is enriched for genes involved in development, ATP synthesis, and the dynein complex, a pattern that may be explained by translational suppression from 30-180 minutes in the familiar condition.

Enriched GO terms in clusters that peak at 30 minutes are quite distinct in the novel and familiar conditions. In both clusters, terms related to development are enriched as well as enzyme-linked receptor activity and transmembrane transport (weakly). In the novel condition, terms relating to adhesion, transcriptional regulation, and the extracellular matrix are also enriched, whereas in the familiar condition, many terms related to the synaptic response are enriched in addition to vesicle transport and the endoplasmic reticulum.

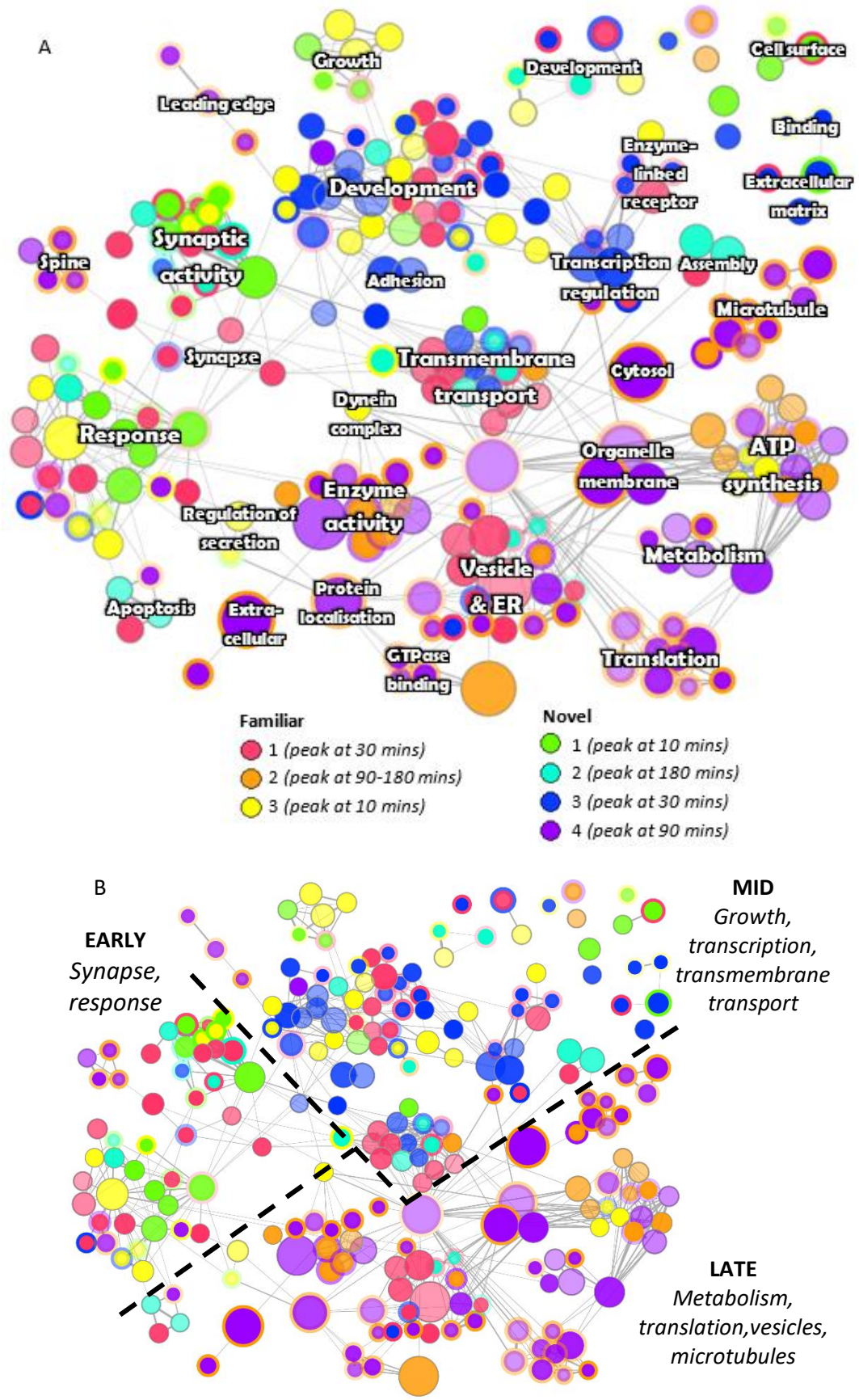


Figure 4.13: Map of enriched gene ontology (GO) terms in clusters of genes with similar expression profiles for novel and familiar conditions. Nodes represent enriched GO terms (p<0.025) in all ontologies and edges represent a similarity coefficient between terms >0.3 (calculated using EnrichmentMap (Merico et al, 2010)

based on gene set overlap). Node colour represents the cluster(s) in which each GO term was enriched (see key). Node size is mapped to the number of significant genes annotated to that term and the largest nodes have >50 significant genes. Node and border transparency are mapped to p-value, such that transparent nodes are less significant (p>0.01). Where nodes have a border, the central colour and node size correspond to the cluster with the lower p-value for that term. A greater significance threshold of p<0.05 was used for secondary enrichments to more clearly visualise cluster overlap. Enrichment of GO terms in more than two clusters is not shown. Edge width is linearly mapped to similarity coefficient so terms with greater overlap are connected by thicker edges. The network was generated in Cytoscape v3.5.1 and the force-directed layout was applied. Clusters were arranged to emphasise groups of interconnected nodes with related functions. A: Manually labelled map of functional groupings (e.g. synaptic activity). All groups of related nodes and some individual nodes with distinct functions are annotated. B: Summarised concept map showing functional groupings of GO terms enriched in clusters of genes with different expression profiles. The "early" group includes primarily terms enriched in clusters F1, F3 and N1 related to synaptic activity and cellular responses; the "mid" group contains primarily terms enriched in F1, F3, N2, and N3 involved in development and growth, transmembrane transport, and transcription, and the "late" group contains primarily terms enriched in F2 and N4 related to translation, metabolism, vesicles and microtubules.

The relatively small cluster of genes with peak expression at 180 minutes in the novel condition is shown in Figure 4.13 to have the closest association with genes that are upregulated at 30 minutes in the familiar condition, consistent with the weaker upregulation of these genes at 30 minutes in the novel condition. Enriched functions in this cluster are related to the regulation of synaptic activity and plasticity, protein complex assembly, transmembrane transport, vesicle membranes, neuron growth, and regulation of apoptosis.

4.3.3.4.1 Functional enrichment in no-image control data clusters

For comparison of the above findings with the no-image control condition, the top five enriched GO terms within each cluster found in the control data are presented in Table 4.6, using a threshold of five annotated genes in the set as above.

Cluster C1, which strongly overlaps with genes upregulated in the novel and familiar conditions at 90 minutes, is most enriched for the cellular compartments cytosol (N=88, $2.6x10^{-7}$) and extracellular exosomes (N=67, $9.8x10^{-6}$). In clusters C2 and C4, in which there is evidence of a translational response at early timepoints, enrichment of GO terms related to receptor activity and stimulus detection is observed, including olfactory receptor activity in C2 (N=7, $5.5 x10^{-9}$) and transmembrane signalling receptor activity in C4 (N=13, 0.00088). In the cluster of genes that show little change in expression over time in the control condition (C3), terms more relevant to downstream effects are enriched, such as skeletal muscle cell differentiation (N=6, p=0.00011), suggesting that these functions are not regulated by the neuronal response to the no-image control condition. These results suggest that temporal changes in gene expression observed in the control condition are associated with synaptic activity , but not with enduring cellular or synaptic modifications .

Table 4.6: Top five enriched GO terms in clusters of genes in the no-image control data

GO	GO ID	Term	Genes in set	P-value
Cluster C1				
CC	GO:0005829	cytosol	88	2.6×10^{-7}
CC	GO:0070062	extracellular exosome	67	9.8×10^{-6}
CC	GO:0005634	nucleus	123	0.00075
BP	GO:0045454	cell redox homeostasis	6	0.00107
CC	GO:0031090	organelle membrane	52	0.00117
Cluster C2				
BP	GO:0050911	detection of chemical stimulus involved in sensory perception of smell	7	5×10^{-9}
MF	GO:0004984	olfactory receptor activity	7	5.5×10^{-9}
MF	GO:0004930	G-protein coupled receptor activity	11	0.000022
CC	GO:0016021	integral component of membrane	50	0.00071
CC	GO:0005615	extracellular space	17	0.00156
Cluster C3				
BP	GO:0035914	skeletal muscle cell differentiation	6	0.00011
BP	GO:0002437	inflammatory response to antigenic stimulus	5	0.00013
BP	GO:0042773	ATP synthesis coupled electron transport	5	0.00069
CC	GO:0016021	integral component of membrane	68	0.0016
BP	GO:0009952	anterior/posterior pattern specification	7	0.00322
Cluster C4				
CC	GO:0031594	neuromuscular junction	6	0.00001
CC	GO:0043195	terminal bouton	6	0.00017
CC	GO:0043235	receptor complex	9	0.00019
CC	GO:0005887	integral component of plasma membrane	15	0.00086
MF	GO:0004888	transmembrane signaling receptor activity	13	0.00088

4.4 Discussion

In this chapter, data were analysed from a multi-factor translatome profiling study of CA1 neurons within a three-hour period following exposure to novel and familiar spatial arrangements of stimuli or a no-image control condition. Multiple approaches to data analysis were applied with the aim of finding reliable differences in the regulation of specific genes and functions across conditions and over time that may contribute to forms of plasticity in CA1 neurons required for associative recognition memory.

Due to the multiple related conditions, the dataset has high information content compared to a hypothetical dataset with the same number of samples but fewer conditions and instead more replicates (Blainey et al, 2014) and interpretation of findings was furthered by comparison of related conditions. Graded increases or decreases in gene expression over time were observed in all conditions, including the unreplicated control data, indicating the sensitivity of the data. As the present study did not include a home-cage control condition, differences due to downregulation in some conditions could not be unambiguously distinguished from differences due to upregulation in others. The present study thus addresses whether gene expression is regulated differently after viewing novel, familiar, or no stimuli and not whether gene expression differs from a static baseline, thereby reflecting the fundamental stochasticity of gene expression (due to random fluctuations in transcription and translation; Raj & van Oudenaarden, 2008) and the continuity of neural processing in vivo.

4.4.1 Methods for differential expression analysis

Multiple pipelines for DE analysis using GLMs with a NB distribution were compared to find the best fitting model for the data, assess the extent of differential expression across conditions, and evaluate the reliability of DEG detection.

4.4.1.1 Global variation across conditions and samples

PCA showed that the effects of the experimental variables on gene expression were relatively weak compared to overall variability, as anticipated, and that the strongest effect was of timepoint, between the 10 and 30-minute and 90 and 180-minute conditions. RUVseq (Risso et al, 2014) was used to compensate for unwanted variation and batch effects and the use of empirically defined negative control genes enabled factors of unwanted variation across all samples to be identified independently of condition, thereby avoiding overfitting and potential bias (Gagnon-Bartsch & Speed, 2012). This method could have been improved by identification of an a priori set of negative control genes using independent samples from similar conditions. Results may also be affected by the partial confounding of experimental variables with batch effects due to incomplete counterbalancing, despite efforts to reduce the influence of technical effects.

Downstream hierarchical clustering of conditions by expression of IEGs and robustly-identified DEGs again showed that effects of condition were small compared to the effects of timepoint. The greatest within-timepoint difference between novel and

familiar conditions was at 180 minutes and may reflect response persistence in the familiar but not the novel condition or a separate wave of gene expression in the novel condition, or simply the low number of replicates. On the other hand, consistent changes in expression over time compared to the other conditions were observed in the no-image control condition even with no replicates. The design enabled the plausibility of observed differences in expression to be evaluated by comparison with related conditions, to identify outliers and differences in expression likely due to noise. Genes with consistent differences in expression across multiple conditions could therefore be robustly identified. For instance, some genes showed increased expression at 10-30 minutes and reduced expression at 90-180 minutes in both novel and familiar conditions and this finding was effectively replicated twelve times, whereas power to detect genes that were differently regulated in only one condition was relatively low.

4.4.1.2 Differential expression pipelines

Differential expression analyses, conducted using edgeR (Robinson et al, 2010; McCarthy et al, 2012) and DESeq2 (Love et al, 2014) to fit NB GLMs, were assessed using the p-value distribution and numbers of identified DEGs. The p-values found by all pipelines were very highly correlated, indicating that the main difference between methods is in the choice of threshold. Very different numbers of DEGs were identified using alternative algorithms in edgeR: the more conservative QL method (Lun et al, 2016) may be insufficiently sensitive for the present exploratory study, whereas the LRT method (McCarthy et al, 2012) is more vulnerable to type 1 error. Promisingly, the robust version of the LRT method in edgeR (Zhou et al, 2014) and the DESeq2 pipeline identified similar and plausible numbers of DEGs (~800), comparable to the number of DEGs identified in a recent TRAP study profiling gene expression 30, 60, and 120 minutes after chemically induced L-LTP in CA3-CA1 mini-slices (899 DEGs; Chen et al, 2017).

By modelling the main sources of variability across all samples, statistically significant DEGs were identified that are likely to be regulated in CA1 following the presentation of visuospatial stimuli. However, this approach failed to accurately model the variability within paired samples and so for exploratory analysis of differences between these conditions, separate models were fitted for novel and familiar samples from each timepoint. Although more experiments are needed to reliably detect differences between novel and familiar conditions, plausible candidate genes and functions that may be differentially regulated during associative recognition memory formation were identified using this approach.

4.4.2 Immediate early gene expression

IEG expression is required for long-term synaptic plasticity and memory consolidation (Tischmeyer & Grimm, 1999; Miyashita et al, 2008; Sun & Lin, 2016) and was used here both as a positive control and for comparison of experimental conditions. The results showed robust differential expression of several IEGs over time.

Several IEGs were more highly expressed at 10-30 minutes compared to 90-180 minutes, including the majority of differentially expressed IEGs, which may be due in part to higher statistical power to detect this pattern of differential expression as discussed

above. In addition, some previously identified primary response genes showed delayed upregulation, such as Bdnf at 30 minutes and Nup98 and Mbnl2 at 90 minutes. Npas4 had a distinctive expression pattern among the IEGs that was characterised by brief and rapid induction, suggesting regulation by different mechanisms. A similar profile of Npas4 expression has been observed previously in vivo, with peak mRNA expression reached within five minutes of CFC compared to thirty minutes for Fos and Arc (Ramamoorthi et al, 2011). Npas4 is expressed in INs as well as PNs and is involved in plasticity of inhibitory connections (Lin et al, 2008; Sun & Lin, 2016).

Groups of primary response genes with similar expression patterns in the present study were also found to be co-regulated in a previous study of IEG expression in cortical neuronal cultures (Saha et al, 2011); genes here upregulated at 10 minutes in novel and familiar conditions, except for Nr4a1, were found by Saha et al to be rapidly transcribed due to promoter-proximal stalling of Pol2. Interestingly, in differential expression analysis of the present dataset, the group of IEGs upregulated at early timepoints were found to have very similar expression patterns to other genes including Notch1, which is known to be positively regulated by Arc (Alberi et al, 2011). This finding could be explained by shared regulatory mechanisms at the translational level.

The majority of IEGs, including Fos and Arc, were not differentially expressed between novel and familiar conditions, despite previous evidence that Fos is more strongly expressed in CA1 neurons following viewing of novel compared to familiar configurations of stimuli (Wan et al, 1999). The difference in the number of Fos-stained nuclei found by IHC in the PV task is relatively small, as shown in Chapter 3 and by Wan et al (1999), and this combined with low numbers of replicates, noise from extraction and sequencing, and pooling across animals may explain the lack of a difference here. On the other hand, consistent with the patterns of IEG expression observed here, Miyashita et al (2008) found that many of the same IEGs were induced 30 minutes post-training on days one and five of learning a spatial water maze task and after reversal learning, but not 180 minutes post-training on day one.

Ultimately, larger experimental groups will be needed to detect robust differential expression between stimulus conditions using paired viewing, yet some of the observed trends in primary response gene expression between conditions suggest potential avenues for further research despite not reaching statistical significance. The relative expression profiles of key IEGs showed that while expression changes were induced in all three conditions, the timing and magnitude varied. Key IEGs (including Arc and Junb) were upregulated in the no-image control condition as has been observed in previous research (Cavallaro et al, 2002; Shires & Aggleton, 2008), but induction was later and of lower magnitude compared to the novel and familiar conditions. Fos expression was unchanged in the control condition. Trends in the expression of three primary response genes, Dusp5, Rasl11a, and Gadd45b, were suggestive of regulation in the familiar condition over time, while expression remained relatively unchanged in the novel condition. These genes are potential candidates for further study and each has been found to contribute to the regulation of gene expression by different mechanisms: Dusp5 is a phosphatase involved in homeostatic regulation of MAPK levels (Iacono et al

2013), while Rasl11a has been found to regulate transcription of rRNA (Pistoni et al, 2010), and Gadd45b may be involved in long-term positive regulation of gene expression by promoting DNA demethylation (Ma et al, 2009). Knockout of Gadd45b has been found to impair hippocampal-dependent LTM (Leach et al., 2012).

4.4.3 Protein-protein interaction network of differentially expressed genes and putative hub genes

Known protein-protein interactions were strongly enriched among DEGs. The network of interactions was used to identify hub proteins that were central to the network and may therefore be particularly important for associative recognition memory formation and interesting candidates for further study. This is a relatively simplistic approach that is limited by the fact that DEGs interact with all proteins present in the cell, not only those that are differentially expressed; a change in the amount of one differentially expressed protein can affect the function of many different constitutively expressed proteins. The present approach is also dependent on databases of previous experimental work and so is necessarily heavily biased towards well-studied genes and pathways, particularly those linked to disease. Nevertheless, the present approach provides a useful starting point to consider known pathways that may contribute to associative recognition memory.

Several of the putative hub genes have been previously found to be central to plasticity processes involved in learning, including Calm2, NMDAR1, and β-catenin, suggesting that this method identifies reliable candidate genes. Calmodulin and NMDARs are critical for the induction of plasticity, while a particularly important role of β-catenin is to mediate adhesion by linking the actin cytoskeleton at both pre- and postsynaptic sites to cadherins, proteins that connect cells via the extracellular matrix (Takeichi & Abe, 2005). Based on these findings, adhesion, DNA repair, and mitochondrial function are predicted to be key molecular processes in the response to paired viewing. Selected hub proteins are discussed in more detail below: Ppp1cb, the small GTPases Rac1 and Cdc42, and cyclin H.

4.4.3.1 Protein phosphatase 1

PP1 was found to be central to the DEG network and is known to mediate LTD (Morishita et al, 2001). Partial inhibition of PP1 shifts synaptic responses to stimulation at a range of frequencies towards potentiation and impairs depotentiation (Jouvenceau et al, 2006). PP1 present at synaptic terminals negatively regulates synaptic strength by reactions such as dephosphorylation of CaMKii at autophosphorylation sites (Lisman & Zhabotinsky, 2001). In the nucleus, PP1 regulates the expression of genes involved in chromatin remodelling and both positive and negative regulators of plasticity (Gräff et al, 2010). In the present study, the catalytic subunit of PP1, Ppp1cb, was differentially expressed. Ppp1cb directs PP1 activity to myosin and troponin (Pereira et al, 2011) and a mutation in Ppp1cb has been associated with intellectual disability (Ma et al, 2016). Overall, these findings imply that passive exposure to visuospatial stimuli negatively effects synaptic strength, in accordance with the findings of Kemp & Manahan-Vaughan (2012).

4.4.3.2 Rho GTPases

The Rho GTPases Cdc42 and Rac1 and the Rho-activated kinase Rock2 were central to the network of DEGs regulated by paired viewing. Rho GTPases each have distinct roles in actin remodelling: Cdc42 and Rac1 regulate the protrusion of filopodia and lamellipodia respectively, whereas Rho is involved in cell contraction and the formation of actin stress fibres (Hall, 1998). At synapses, Rho GTPases have been found to regulate the density, morphology, and dynamics of spines and AMPAR clustering (Tashiro & Yuste, 2004; Wiens et al, 2005; Haditsch et al, 2009). Rho GTPases also regulate gene expression via signalling pathways (Coso et al, 1995). Interestingly, a recent study found that synaptic Rho GTPases are activated with different dynamics: Cdc42 is activated by relatively weak stimulation and restricted to individual spines, whereas Rac1 and Rho diffuse to nearby spines (Hedrick et al, 2016).

In behavioural studies, knockout of Rac1 has been found to impair working memory (Haditsch et al, 2009), whereas knockout of Cdc42 impaired remote memory (Kim et al, 2014). Cdc42 and Rac1 have also been found to mediate forgetting of different forms of memory. In Drosophila, Cdc42 has been found to regulate forgetting of anaesthesia-resistant memory (Zhang et al, 2016) and Rac1 has been found to mediate forgetting of anaesthesia-sensitive memory (Shuai et al, 2010), as well as NOR memory in mice (Liu et al, 2016). Based on these findings, it has been proposed that Cdc42 is required for the generation of new spines and modulates labile forms of memory, whereas Rac1 is required for the stabilisation of structural plasticity and modulates consolidated memories (Haditsch et al, 2009; Zhang et al, 2016).

4.4.3.3 Cyclins

Cyclins have been found to regulate synaptic plasticity and memory (Frank & Tsai, 2009). Cyclin Y, for example, has been shown to negatively regulate LTP by inhibition of AMPAR exocytosis (Cho et al, 2015a). Conversely, LTP and spine volume are reduced in knockout mice deficient in cyclin E, which regulates Cdk5 (Odajima et al, 2011). Analysis of the DEG network in the present study suggests that cyclin H may participate in associative recognition memory in CA1. Cyclin H activates Cdk7 (Tang & Wang, 1996) and Cdk7 activates the basal transcription factor TfIIb, Cdk5, and other Cdks (Rosales et al, 2003). Increased Cdk5 expression in the hippocampus has been shown to be required for associative learning (Fischer et al, 2002). Multiple roles for Cdk5 in neurons and synaptic plasticity have been identified, including regulation of NMDAR currents (Angelo et al, 2006; Hawasli et al, 2007), suggesting that Cdk5 may contribute to metaplasticity. Further research is needed to determine the contribution of cyclin H to hippocampal-dependent learning.

4.4.4 Clustering of co-expressed genes

Genes known to have related functions, such as genes belonging to the PSD, were found to have highly correlated expression profiles and co-expressed genes identified by clustering were found to be enriched for functions relevant to synaptic plasticity and learning. Clusters of varying coherence, defined by different parameters and clustering algorithms, may correspond to different levels of regulation and further research is

necessary to understand this relationship. For example, groups of genes found to have highly similar expression profiles and closely related functions may represent co-regulated units (such as synexpression units described by Niehrs & Pollet, 1999), whereas the much larger k-means clusters may correspond to the output of the averaged activity of a combination of key upstream regulators, with sub-groups of genes and individual genes being subject to additional levels of regulation.

K-means clustering of DEGs was applied to study broad differences in expression between conditions, after standardising expression profiles to focus on proportional and relative changes in expression, irrespective of gene abundance. Four clusters of genes were identified in the novel and control conditions and three clusters were identified in the familiar condition using these methods. In the no-image control condition, clusters were differentially regulated only at early timepoints (10-30 minutes), suggesting that expression changes were short-lived, whereas in the novel and familiar conditions, clusters of genes were differentially regulated at all timepoints. In the novel condition, a different cluster of genes was most strongly expressed at each timepoint, while in the familiar condition, expression at 90 minutes was sustained at 180 minutes.

4.4.4.1 Differential timing of translational responses across conditions

All conditions included a cluster of genes with peak expression at 10 minutes followed by a gradual reduction in expression and plateau (a profile that may be consistent with early upregulation or later repression) and a cluster with peak expression at 30 minutes. The greatest gene set overlap between novel and familiar conditions among these clusters was between those with similar expression profiles, though many genes were assigned to clusters with different expression profiles in the two conditions. In the no-image control condition in contrast, clusters with a 10-minute peak in the experimental conditions most overlapped with the cluster strongly upregulated at 30 minutes, while the clusters with a 30-minute peak in the experimental conditions most overlapped with a cluster that showed little change in expression over time. These expression patterns suggest that the former group of genes may have repressive or homeostatic effects on plasticity (Abel & Kandel, 1998), while the latter are more likely to be required for Hebbian forms of plasticity. Genes with peak expression at 10 minutes in the control condition did not strongly overlap with any novel or familiar cluster and may be involved in memory suppression. A large group of genes were upregulated at 90 minutes in the novel and familiar conditions and showed little change in expression over time in the no-image control conditions, though expression profiles indicate potential mild downregulation at early timepoints in all conditions. This expression pattern suggests that these genes may be involved in structural changes at the synapse. The cluster of genes in the novel condition with fluctuating expression and a 180-minute peak did not strongly overlap with any other cluster.

Analysis of expression profiles suggests that translational responses are initially induced in all three conditions in response to activity, but response persistence may depend on stimulus salience. The translational response may decay after 30 minutes in the no-image condition, while differential expression of downstream or late-response genes may be induced (at 90 minutes) in both the novel and familiar conditions and sustained

(at 180 minutes) only in the familiar condition. A different set of genes may be regulated in the novel condition at 180 minutes, but more replicates are needed to validate this finding. These findings imply that learning is occurring in the familiar condition and contradict previous studies that have found that gene expression is most strongly regulated in response to novel stimuli and habituates over repeated exposures (Anokhin & Rose, 1991; Nikolaev et al, 1992; Bertaina-Anglade et al, 2000). Learning may be slow in the present study due to the passive nature of the paired viewing paradigm, the absence of reinforcement, and the very short duration of exposures. The amount of exposure to a stimulus required for the habituation of gene expression is unclear and likely to be modulated by stimulus saliency and attention. In previous studies for example, habituation of IEG expression has been observed following three consecutive days of approximately ten minutes of training on spatial maze tasks (Papa et al, 1993; Cavallaro et al, 2002) or during one long exposure to a novel stimulus lasting 15-30 minutes (Mello et al, 1995; Kemp & Manahan-Vaughan, 2012). Another factor that may account for the contradictory results is that the familiar image set was last presented to subjects three hours prior to the test session in the previous experiments, whereas here the previous training session was ~21 hours prior. Different mechanisms underlie the persistence of memory at 24 hours compared to three hours (Kandel et al, 2014). Cross-communication between hemispheres may also contribute to differences in the timing of expression changes between novel and familiar conditions (Gazzaniga, 1966).

Overall, these results suggest that different conditions in the paired viewing protocol affect the expression of shared sets of proteins with different dynamics. As RTFs are known to have combinatorial effects on regulation (Joo et al, 2016), these expression profiles may result from differential activation of overlapping upstream regulatory pathways (Coba et al, 2008; Jain & Bhalla, 2014).

4.4.4.2 Functional enrichment profiles

Differences in functional enrichment over time were found in the novel and familiar conditions by mapping GO terms that were significantly enriched in different clusters by gene set overlap. As individual genes are involved in multiple functions (an average of ten according to one estimate; Miklos & Rubin, 1996), the same protein may contribute to different processes depending on cellular context (D'haeseleer et a, 2000) and here may have different functional roles in different conditions. For example, Arc expression has been shown to be required for several different forms of plasticity, including homeostatic scaling (Shepherd et al, 2006), L-LTP (Guzowski et al, 2000; Plath et al, 2006), and mGluR-dependent LTD (Plath et al, 2006; Waung et al, 2008).

Clusters with peak expression at 10 minutes and reduced expression at later timepoints in both the novel and familiar conditions were enriched for terms related to the synapse, receptor activity, and response to stimuli. In the familiar condition, terms related to development and ATP synthesis were additionally enriched among genes with this expression profile. Synaptic and response terms were also enriched in the cluster of genes with peak expression at 30 minutes in the familiar condition and clusters of genes with increased expression at 10-30 minutes in the control condition.

In both novel and familiar conditions, clusters with peak expression at 30 minutes were enriched for terms related to transmembrane transport, enzyme-linked receptor activity, and development. These genes may be involved in prolonging activity-dependent responses and initiating downstream changes in expression. In the familiar condition, terms related to vesicle transport and the endoplasmic reticulum were also enriched, whereas in the novel condition, terms relating to transcriptional regulation and adhesion (including vascular development and the extracellular matrix) were enriched. These findings suggest that the response to viewing novel arrangements may involve the formation of new synaptic contacts, whereas the strengthening of associations over repeated presentations may involve receptor trafficking.

At 90 minutes in both experimental conditions, functions related to the structural remodelling of synapses were upregulated, including spine formation, microtubule activity, and actin remodelling, suggesting that structural plasticity occurs following exposure to both novel and familiar stimuli. Genes involved in translation, metabolism, and GTPase binding were enriched primarily in the novel condition at this timepoint and ATP synthesis was enriched in the familiar condition.

In the cluster of genes with peak expression at 180 minutes in the novel condition, enriched functions included transmembrane transport, protein complex assembly, and regulation of apoptosis. Enrichment of genes involved in the negative regulation of apoptosis has been reported previously following induction of plasticity and may promote neuron survival (Iacono et al, 2013). However, more replicates are needed to investigate differences between conditions at 180 minutes.

GO analysis of the control condition showed that clusters that showed upregulated expression at early timepoints were enriched for genes linked to receptor activity, whereas genes linked to structural changes at the synapse and development were enriched in clusters that showed little change in expression over time. These findings support an interpretation of gene expression in the control condition as being related to the response to stimuli but not to downstream changes in expression.

4.4.5 Summary

The present study shows that protein synthesis is regulated in CA1 neurons following exposure to both novel and familiar visuospatial stimuli, as well as in the no-image control condition. Changes in gene expression were broadly similar in the novel and familiar conditions and differed in the no-image condition. There was insufficient power to detect differential gene expression between the novel and familiar conditions. In the no-image control condition, differential gene expression was observed only at early timepoints and was enriched for genes involved in receptor activity and the synaptic response. These findings suggest that long-term changes are unlikely to be induced in the no-image condition and highlight the importance of a behaviourally-matched control condition to distinguish between gene expression induced by the manipulation of interest and gene expression induced by other aspects of behavioural tasks.

In both novel and familiar conditions, IEGs and genes known to play important roles in memory formation were differentially expressed. Overall, genes involved in the synaptic response were upregulated in the translatome at early timepoints and genes involved in synaptic remodelling and metabolism were upregulated at later timepoints. PP1 and the Rho GTPases Rac1 and Cdc42 are predicted to be key regulators of the response to paired viewing. The present findings highlight the need for further investigation into the molecular mechanisms that underlie incremental and passive forms of learning.

4.5 Supplemental information

Table 4.7: Differentially expressed genes detected by all pipelines

Gene ID	Full name	logCPM	FDR
Npas4 *(ENSRNOG00000020009)*	neuronal PAS domain protein 4	3.35	3.22E-14
Arc *(ENSRNOG00000043465)*	activity-regulated cytoskeleton-associated protein	7.13	3.22E-14
Notch1 *(ENSRNOG00000019322)*	notch 1	4.07	3.86E-08
Dusp1 *(ENSRNOG00000003977)*	dual specificity phosphatase 1	5.16	1.37E-07
Tmem69 *(ENSRNOG00000029152)*	transmembrane protein 69	5.08	1.10E-05
Prmt3 *(ENSRNOG00000014829)*	protein arginine methyltransferase 3	5.64	2.02E-05
Gdf11 *(ENSRNOG00000007610)*	growth differentiation factor 11	4.25	3.48E-05
Snx17 *(ENSRNOG00000026884)*	sorting nexin 17	5.80	3.48E-05
Fam174a *(ENSRNOG00000019087)*	family with sequence similarity 174, member A	3.49	3.93E-05
Grik2 *(ENSRNOG00000000368)*	glutamate ionotropic receptor kainate type subunit 2	5.77	5.74E-05
Nr4a1 *(ENSRNOG00000007607)*	nuclear receptor subfamily 4, group A, member 1	7.93	6.96E-05
Hspa9 *(ENSRNOG00000019525)*	heat shock protein family A member 9	7.53	1.18E-04
Gca *(ENSRNOG00000007359)*	grancalcin	5.22	1.28E-04
Usp45 *(ENSRNOG00000008688)*	ubiquitin specific peptidase 45	6.52	1.89E-04
Taf13	TATA-box binding protein	3.34	1.89E-04

(ENSRNOG00000020315)	associated factor 13		
Cryl1 *(ENSRNOG00000008989)*	crystallin, lambda 1	4.77	2.19E-04
Ptn *(ENSRNOG00000011946)*	pleiotrophin	6.53	2.99E-04
Egr2 *(ENSRNOG00000000640)*	early growth response 2	3.80	3.42E-04
Acot9 *(ENSRNOG00000003782)*	acyl-CoA thioesterase 9	5.05	3.63E-04
Junb *(ENSRNOG00000042838)*	JunB proto-oncogene, AP-1 transcription factor subunit	4.23	4.83E-04
Anp32a *(ENSRNOG00000014846)*	acidic nuclear phosphoprotein 32 family member A	4.96	5.33E-04
Cenpb *(ENSRNOG00000057284)*	centromere protein B	3.56	5.33E-04
Btg2 *(ENSRNOG00000003300)*	BTG anti-proliferation factor 2	4.37	5.74E-04
Mrpl35 *(ENSRNOG00000008546)*	mitochondrial ribosomal protein L35	4.85	5.74E-04
Cct8 *(ENSRNOG00000001592)*	chaperonin containing TCP1 subunit 8	7.04	6.38E-04
Eif1 *(ENSRNOG00000033765)*	eukaryotic translation initiation factor 1	7.94	7.46E-04
Psat1 *(ENSRNOG00000013971)*	phosphoserine aminotransferase 1	6.60	8.11E-04
Tpp2 *(ENSRNOG00000011194)*	tripeptidyl peptidase 2	5.75	8.25E-04
Picalm *(ENSRNOG00000018322)*	phosphatidylinositol binding clathrin assembly protein	7.77	8.37E-04
Gk *(ENSRNOG00000034116)*	glycerol kinase	4.21	8.61E-04
Actr2 *(ENSRNOG00000004959)*	ARP2 actin related protein 2 homologue	10.10	8.95E-04

Lmtk3 *(ENSRNOG00000021048)*	lemur tyrosine kinase 3	3.19	8.95E-04
Ap1s2 *(ENSRNOG00000038686)*	adaptor-related protein complex 1, sigma 2 subunit	4.74	8.95E-04
Mrfap1 *(ENSRNOG00000055319)*	Morf4 family associated protein 1	8.62	8.95E-04
AABR07009834.1 *(ENSRNOG00000061630)*	*Not annotated*	3.76	8.95E-04
Nap1l3 *(ENSRNOG00000029087)*	nucleosome assembly protein 1-like 3	6.23	9.01E-04
Hnrnpk *(ENSRNOG00000019113)*	heterogeneous nuclear ribonucleoprotein K	6.83	9.20E-04
Ubash3b *(ENSRNOG00000008187)*	ubiquitin associated and SH3 domain containing, B	6.39	1.13E-03
Scg5 *(ENSRNOG00000007542)*	secretogranin V	8.19	1.23E-03
Plxna1 *(ENSRNOG00000017003)*	plexin A1	7.60	1.23E-03
Arhgef6 *(ENSRNOG00000000869)*	Rac/Cdc42 guanine nucleotide exchange factor 6	3.72	1.23E-03
Drap1 *(ENSRNOG00000020527)*	Dr1 associated protein 1	8.18	1.46E-03
Eif3m *(ENSRNOG00000012738)*	eukaryotic translation initiation factor 3, subunit M	6.01	1.50E-03
Dnhd1 *(ENSRNOG00000051291)*	dynein heavy chain domain 1	3.02	1.50E-03
Cops2 *(ENSRNOG00000008744)*	COP9 signalosome subunit 2	7.06	1.55E-03
Asxl1 *(ENSRNOG00000061603)*	additional sex combs like 1	5.11	2.25E-03
Extl1 *(ENSRNOG00000016776)*	exostosin-like glycosyltransferase 1	5.28	3.23E-03
Cdc42 *(ENSRNOG00000013536)*	cell division cycle 42	8.82	3.61E-03

Cse1l *(ENSRNOG00000007665)*	chromosome segregation 1 like	6.21	3.97E-03
Arl8b *(ENSRNOG00000055860)*	ADP-ribosylation factor like GTPase 8B	8.46	3.98E-03
Rab6a *(ENSRNOG00000018176)*	RAB6A, member RAS oncogene family	9.24	4.24E-03
P4ha1 *(ENSRNOG00000050655)*	prolyl 4-hydroxylase subunit alpha 1	5.48	8.24E-03
Dab1 *(ENSRNOG00000007410)*	DAB1, reelin adaptor protein	8.26	1.75E-02
Aak1 *(ENSRNOG00000018317)*	AP2 associated kinase 1	10.37	2.38E-02

Chapter 5 General Discussion

This thesis describes a study utilising TRAP (Heiman et al, 2008; Doyle et al, 2008) and paired viewing of arrangements (Wan et al, 1999) to profile the translatome of rat CA1 neurons during associative recognition memory formation, as well as the optimisation of methods for this purpose, including the development of viruses for TRAP.

5.1 TRAP viruses

In Chapter 2, AAV and lentiviruses expressing EGFP-L10a under the control of a CMV promoter were developed and the AAV was found to produce very strong and relatively specific expression of the transgene in CA1 neurons, while expression of the lentivirus was specific to pyramidal neurons in hippocampal cultures. A subregion-selective dissection was used to compensate for potential spread of the virus outside CA1. TRAP samples were enriched for markers of CA1 and pyramidal neurons compared to other cell types. Transcripts from all cellular compartments were identified, including the nucleus and postsynaptic density. Expression of transcripts specific to astrocytes and interneurons was also observed at a lower level and some of these genes were differentially expressed between novel and familiar paired samples. In future experiments, CA1 neurons or other cell types may be targeted more precisely using a recently developed Cre-dependent TRAP virus (Nectow et al, 2017), in combination with a virus using a larger promoter, to prevent the need for a subregion-selective dissection.

5.2 Learning in the paired viewing paradigm

Fos expression was found to be significantly higher in the novel than the familiar condition in Chapter 3, when expression of Fos protein was measured by IHC (replicating Wan et al, 1999), but not in Chapter 4, when expression of ribosome-associated Fos mRNA was measured using RNAseq. This discrepancy may be due to a small effect being obscured by noise in the sequencing data or differences in the timing of peak differential expression of Fos ribosome-bound mRNA compared to Fos protein as the former may have occurred in-between the 30- and 90-minute timepoints.

In the translatome profiling data, known key mediators of plasticity processes were differentially expressed and similar patterns of differential expression were observed in the novel and familiar samples that differed in the no-image control condition (see below). These findings imply that learning is occurring in both the novel and familiar conditions. In any behavioural study, it is difficult to be sure of what subjects are truly learning and this may vary between animals. The mechanisms of associative learning, as advanced by Rescorla (1967), are complicated and off-target effects can be difficult to rule out (Nikolaev et al, 1992). In the no-image control condition of the present study for example, the absence of visual stimuli may violate expectations and lead to reconsolidation or extinction processes requiring protein synthesis (Pedreira et al, 2004). As highlighted by Barto et al (2013), an unanticipated experience causes surprise, which differs from novelty caused by a stimulus having not been experienced previously. Gene

expression in each of the three conditions in the present study may correspond to a range of processes, including memory formation, consolidation, retrieval, reconsolidation, and extinction.

The present study does not address whether long-term memory and long-lasting neuronal plasticity were induced in response to the paired viewing paradigm. In a study using similar behavioural apparatus and a single long (15-minute) image presentation, STD was elicited in CA1 in response to novelty, whereas the addition of low-frequency stimulation was required to induce LTD (Kemp & Manahan-Vaughan, 2012). A possible explanation for the present findings is that multiple presentations of stimuli were required for LTM formation due to the short duration and passive nature of stimulus exposures. Most studies of gene expression during learning have used paradigms to induce one-shot learning (e.g. CFC) and few studies have investigated gene expression following repeated training sessions using incremental learning paradigms (e.g. spatial maze training). One study using a spatial water maze task found that differential gene expression was highest after three days of training (Cavallaro et al, 2002), suggesting that gene expression is not always strongest in response to novel stimuli.

The present findings suggest a need for more studies into the molecular and synaptic mechanisms that underlie both incremental and passive forms of learning, which may may hold greater relevance for memories formed from everyday experience compared to learning under highly rewarding or strongly aversive conditions (Fusi et al, 2005).

5.3 Differential expression following paired viewing

The translatome profiling data was analysed in Chapter 4 using a variety of approaches combining GLM-based analyses of differential expression, analysis of protein-protein interaction networks, clustering of gene expression profiles, and analysis of enriched GO terms. IEGs such as Fos and Arc, known to be important for LTM formation (Cole et al, 1989; Guzowski et al, 1999), were found to be differentially expressed over time in the three hours following paired viewing of arrangements but were not differentially expressed between novel and familiar conditions. More replicates and/or an alternative design are needed to probe differential expression between novel and familiar conditions.

Analysis of differential expression across all conditions revealed a network of DEGs with known protein-protein interactions that may be involved in the formation of visual associative recognition memories. Hub genes were identified based on network topology that are known to be key regulators of plasticity, including Rac1, Cdc42, the β subunit of PP1, and β-catenin (Hall, 1998; Morishita et al, 2001; Jouvenceau et al, 2006; Haditsch et al, 2009; Maguschak & Ressler, 2012). Hub genes were also identified that are candidates of interest for further study, including Phka1, cyclin H, Rfc4, and Gnal. A number of the DEGs and enriched GO terms identified in the present study are known to be associated with synaptic depression, in particular PP1 and genes involved in vesicle endocytosis. This result is supported by the previous work of Manahan-Vaughan and colleagues, finding that LTD rather than LTP is involved in the encoding of a form of

associative recognition memory, object-in-place memory (Manahan-Vaughan and Braunewell, 1999; Kemp & Manahan-Vaughan, 2004). DEGs linked to structural plasticity, including Actr2 and Sept2, were also identified and many of these genes were upregulated at 90 minutes in the novel and familiar conditions relative to other timepoints and the no-image control condition.

DEGs were clustered to identify groups of genes with similar expression profiles over time within each stimulus condition and thus provide an overview of broad differences in expression between these conditions. The results of clustering showed that overall expression profiles were similar between the novel and familiar conditions, but analysis of enriched GO terms within genes assigned to clusters suggested that certain functions may be differentially regulated in the two conditions, including adhesion in the novel condition and vesicle transport in the familiar condition. To speculate, these findings could reflect the formation of new synaptic contacts in the novel condition and changes to the strength of existing synapses in the familiar condition. However, further research is necessary to validate these results. Differential expression was also observed in the no-image control condition at the two earlier timepoints (10 and 30 minutes) but not at later timepoints, suggesting that long term changes do not occur in this condition and illustrating the importance of matched control conditions in studies of learning and memory.

5.4 Limitations

There were several limitations in the present study that must be addressed in future work and validation studies. Firstly, although the within-subjects design had the advantage of controlling for most sources of unwanted variability, the size of the effect of the presentation of novel compared to familiar stimuli on gene expression using the paired viewing procedure is relatively small and there was insufficient power to detect differences between the novel and familiar conditions in the present study. There is also no behavioural measure to ensure that subjects are learning the associations between stimuli. In future studies, the object-in-place task or the bow-tie maze (Albasser et al, 2010) may be used to incorporate explicit measures of learning and to potentially increase the effect size by eliminating cross-communication between hemispheres as a confounding factor. On the other hand, increased behavioural variability may be an issue for gene expression profiling using these tasks. The complexity of associative recognition memory may necessitate longer than usual training schedules to observe a robust effect on genome-wide gene expression.

Another limitation of the present study is that the quality of TRAP-generated RNA samples was affected by RNA degradation and contamination by MLV-related sequences, thus reducing the number of reads in each sample and leading to the loss of samples. The protocol could be improved by identifying and removing the source of MLV sequence contamination and by reducing RNA degradation by homogenising tissue from each subject immediately and by using an RNase-free hood. Improvements to the RNA isolation procedure may also remove the need for an amplification step, which is an additional source of noise in the present study. Alternatively, bias due to non-linear PCR

amplification may be overcome by barcoding samples using unique molecular identifiers to enable natural duplication to be distinguished from duplication due to PCR (Parekh et al, 2016).

The loss of samples due to RNA degradation also led to there being unequal numbers of replicates in the different conditions and to samples belonging to different conditions being unevenly distributed across sequencing batches. These factors reduced the power to detect differential expression between conditions and increased the influence of batch effects, limiting the scope of the analysis. The study would also have been improved by the inclusion of a home-cage control condition as a static baseline. Without this condition, it was not always possible to unambiguously distinguish between upregulation in one condition and downregulation in others.

Analysis of the dataset could be extended using weighted gene co-expression network analysis (WGCNA; Zhang & Horvath, 2005), an alternative approach to the identification of clusters of co-expressed genes and hub genes. WGCNA is based on analysis of pairwise correlations between all expressed genes and so the method is not limited to well-studied genes or affected by the setting of arbitrary thresholds for differential expression, which are limitations of the present analysis.

Overall, more replicates are needed to robustly detect differentially expressed genes between the novel and familiar conditions and analyse the effect of the interaction between condition and timepoint.

5.5 Validation of findings and future work

Follow-up experiments are necessary to validate the findings of the present exploratory study. Preliminary validation experiments can be conducted using qPCR on the existing samples to validate the differential expression of selected genes, such as the putative hub genes discussed above or the gene that was differentially expressed between the novel and familiar conditions at 10 minutes, Taf13. To determine whether differential expression is caused by regulation at the level of transcription or of translation, qPCR may also be conducted on the input RNA samples from the present study, which were collected and stored at -80°C.

Genes that are found to be differentially expressed in the original samples must be validated in independent samples obtained following paired viewing or a related behavioural paradigm, such as the object-in-place task. Expression profiling may be conducted using qPCR, RNAseq, or quantitative proteomics. By using protocols with a behavioural outcome measure, gene expression can be correlated with learning (e.g. Guzowski et al, 2001) as differences in gene expression between subjects may be explained by differences in learning rate.

The functional importance of genes of interest can be investigated by inhibiting gene expression using siRNAs, transgenic knock-out animals, or pharmacological approaches. Interesting candidates include the regulator of LTD, PP1β, and the novel targets cyclin H, Phka1, and Rfc4. The activity of these genes of interest may also be pharmacologically

increased prior to a behavioural task, to determine whether there are bidirectional effects on learning. The present findings suggest a multitude of avenues to further investigate the molecular mechanisms involved in associative recognition memory formation.

5.6 Quantitative methods to discover molecular mechanisms underlying learning

The present findings indicate that changes in gene expression occurring in response to subtle behavioural manipulations can be detected using translatome profiling as, although the magnitude of differential expression was low, the identification of functionally related genes with similar dynamical expression patterns was robust and the findings were well supported by the literature. The analyses are consistent with a view of plasticity in which the expression of large sets of plasticity-related genes is regulated differently in response to different stimuli leading to different outcomes. This view is supported by computational modelling studies illustrating the complexity of gene expression regulated by kinase networks in response to different stimuli (Coba et al, 2008; Jain & Bhalla, 2014) and by experimental work finding that genes with opposing effects on plasticity are co-regulated (Iacono et al, 2013), that overlapping sets of genes are expressed during different memory processes (Barnes et al, 2012; Poplawski et al, 2014), and that a balance of different forms of plasticity is required for memory precision and robustness (Malenka & Bear, 2004; Kemp & Manahan-Vaughan, 2007).

Accordingly, studies focusing on quantitative differences in the magnitude and timing of expression of a relatively large number of genes, as opposed to approaches focused on discrete differences in expression, may better reflect the reality that learning about the world, the modification of neural networks by changes in synaptic and neuronal properties, and the regulation of gene expression are all continuously occurring phenomena. The development of reliable methods to integrate datasets from different studies is likely to be key to studying the dynamical properties of the gene expression networks involved in plasticity processes. For example, a recently developed tool for aggregation of gene expression data, GEN3VA, was successfully used to identify molecular signatures of aging (Gundersen et al, 2016). Being one of the most studied regions of the brain, the CA1 is particularly suitable for modelling studies as detailed information is available regarding its cellular composition and connectivity (Megias, 2001; Bezaire & Soltesz, 2013), key signalling pathways (Ma'ayan et al, 2005), and plasticity mechanisms (Edelmann et al, 2017).

Computational analyses can also be applied to reconstruct underlying regulatory networks, such as by identifying a minimal set of inputs that can define the output for each gene at a subsequent timepoint within large datasets (Akutsu et al, 2000). One approach that could be applied to the present dataset is to link DEGs and clusters of co-regulated genes to specific regulators by finding enriched motifs in DEG promoters that correspond to specific transcription factors.

5.7 Summary

The present study profiling the translatome of CA1 neurons following paired viewing of novel and familiar image-arrangements identified a large number of genes that were differentially expressed in response to paired viewing. The DEGs included IEGs, genes involved in structural plasticity at synapses, and regulators of gene expression. Robust differential expression was observed between timepoints in all conditions, but more replicates and/or an alternative design is needed to probe differences between the novel, familiar, and no-image control conditions. The findings further the understanding of the networks of genes that are differentially expressed during associative recognition memory formation and identify a range of targets for further study.

The present study illustrates the complexity of responses to stimuli in CA1 cells, in line with the biology of plasticity which involves the integration of the activity of many different receptors, signalling pathways, transcription factors, and plasticity-related effector proteins, as well as the activity of other cell types such as interneurons and astrocytes, to express a variety of long term changes to synaptic and cellular properties. It is clear that plasticity involves an extremely broad range of phenomena and advances in large-scale computational modelling are likely to be key to unravelling the complex interacting signalling pathways and effects on gene expression that underlie different long-term responses to stimuli.

References

Abbas, A. K. (2013). Evidence for constitutive protein synthesis in hippocampal LTP stabilization. *Neuroscience*, *246*, 301-311.

Abel, T., Martin, K. C., Bartsch, D., & Kandel, E. R. (1998). Memory suppressor genes: inhibitory constraints on the storage of long-term memory. *Science*, *279*(5349), 338-341.

Abraham, W. C. (2008). Metaplasticity: tuning synapses and networks for plasticity. *Nature Reviews Neuroscience*, *9*(5), 387.

Abraham, W. C., & Bear, M. F. (1996). Metaplasticity: the plasticity of synaptic plasticity. *Trends in neurosciences*, *19*(4), 126-130.

Abudula, A., Grabbe, A., Brechmann, M., Polaschegg, C., Herrmann, N., Goldbeck, I., Dittmann, K., & Wienands, J. (2007). SLP-65 signal transduction requires Src homology 2 domain-mediated membrane anchoring and a kinase-independent adaptor function of Syk. *Journal of Biological Chemistry*, *282*(39), 29059-29066.

Adams, J. P., & Dudek, S. M. (2005). Late-phase long-term potentiation: getting to the nucleus. *Nature Reviews Neuroscience*, *6*(9), 737.

Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D. S., Busby, M. A., Berlin, A. M., Sivachenko, A., Thompson, D.A., Wysoker, A., Fennell, T., & Gnirke, A. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nature methods*, *10*(7), 623.

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., & Grüning, B. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic acids research, gkw343.

Aggleton, J. P., Brown, M. W., & Albasser, M. M. (2012). Contrasting brain activity patterns for item recognition memory and associative recognition memory: insights from immediate-early gene functional imaging.*Neuropsychologia*, *50*(13), 3141-3155.

Ainge, J. A., Heron-Maxwell, C., Theofilas, P., Wright, P., de Hoz, L., & Wood, E. R. (2006). The role of the hippocampus in object recognition in rats: examination of the influence of task parameters and lesion size. *Behavioural brain research*, *167*(1), 183-195.

Ainsley, J. A., Drane, L., Jacobs, J., Kittelberger, K. A., & Reijmers, L. G. (2014). Functionally diverse dendritic mRNAs rapidly associate with ribosomes following a novel experience. *Nature communications*, *5*.

Aken, B. L., Achuthan, P., Akanni, W., Amode, M. R., Bernsdorff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., & Gil, L. (2017). Ensembl 2017. *Nucleic acids research*, *45*(D1), D635-D642.

Akli, S., Caillaud, C., Vigne, E., Stratford-Perricaudet, L. D., Poenaru, L., Perricaudet, M., Kahn, A., & Peschanski, M. R. (1993). Transfer of a foreign gene into the brain using adenovirus vectors. *Nature genetics*, *3*(3), 224.

Akutsu, T., Miyano, S., & Kuhara, S. (2000). Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics*, *16*(8), 727-734.

Albasser, M. M., Amin, E., Lin, T. C. E., Iordanova, M. D., & Aggleton, J. P. (2012). Evidence that the rat hippocampus has contrasting roles in object recognition memory and object recency memory. *Behavioral neuroscience*, *126*(5), 659.

Albasser, M. M., Chapman, R. J., Amin, E., Iordanova, M. D., Vann, S. D., & Aggleton, J. P. (2010a). New behavioral protocols to extend our knowledge of rodent object recognition memory. *Learning & Memory*, *17*(8), 407-419.

Albasser, M. M., Poirier, G. L., & Aggleton, J. P. (2010b). Qualitatively different modes of perirhinal–hippocampal engagement when rats explore novel vs. familiar objects as revealed by c-Fos imaging. *European Journal of Neuroscience*, *31*(1), 134-147.

Alberi, L., Liu, S., Wang, Y., Badie, R., Smith-Hicks, C., Wu, J., Pierfelice, T.J., Abazyan, B., Mattson, M.P., Kuhl, D., & Pletnikov, M. (2011). Activity-induced Notch signaling in neurons requires Arc/Arg3. 1 and is essential for synaptic plasticity in hippocampal networks. *Neuron*, *69*(3), 437-444.

Alberini, C. M. (2009). Transcription factors in long-term memory and synaptic plasticity. *Physiological reviews*, *89*(1), 121-145.

Alexa, A., & Rahnenführer, J. (2010). topGO: enrichment analysis for gene ontology. *R package version*, *2*(0).

Alexa, A., Rahnenführer, J., & Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, *22*(13), 1600-1607.

Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, *8*(6), 450-461.

Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., & Levine, A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences*, *96*(12), 6745-6750.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res. 25*:3389-3402.

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome biology*, *11*(10), R106.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166-169.

Andersson, M., Blomstrand, F., & Hanse, E. (2007). Astrocytes play a critical role in transient heterosynaptic depression in the rat hippocampal CA1 region. *The Journal of physiology*, *585*(Pt 3), 843-852.

Andrews, S. (2010). FastQC. *A quality control tool for high throughput sequence data*. Babraham Institute, Cambridge, United Kingdom.

Angelo, M., Plattner, F., & Giese, K. P. (2006). Cyclin-dependent kinase 5 in synaptic plasticity, learning and memory. *Journal of neurochemistry*, *99*(2), 353-370.

Aniento, F., Emans, N., Griffiths, G., & Gruenberg, J. (1993). Cytoplasmic dynein-dependent vesicular transport from early to late endosomes. *The Journal of cell biology*, *123*(6), 1373-1387.

Anliker, B., Abel, T., Kneissl, S., Hlavaty, J., Caputi, A., Brynza, J., Schneider, I.C., Münch, R.C., Petznek, H., Kontermann, R.E., & Koehl, U. (2010). Specific gene transfer to neurons, endothelial cells and hematopoietic progenitors with lentiviral vectors. *Nature methods*, *7*(11), 929.

Anokhin, K. V., & Rose, S. P. (1991). Learning-induced Increase of Immediate Early Gene Messenger RNA in the Chick Forebrain. *European Journal of Neuroscience*, *3*(2), 162-167.

Antion, M. D., Hou, L., Wong, H., Hoeffer, C. A., & Klann, E. (2008). mGluR-dependent long-term depression is associated with increased phosphorylation of S6 and synthesis of elongation factor 1A but remains expressed in S6K-deficient mice. *Molecular and cellular biology*, *28*(9), 2996-3007.

Araque, A., Parpura, V., Sanzgiri, R. P., & Haydon, P. G. (1999). Tripartite synapses: glia, the unacknowledged partner. *Trends in neurosciences*, *22*(5), 208-215.

Arava, Y., Wang, Y., Storey, J. D., Liu, C. L., Brown, P. O., & Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, *100*(7), 3889-3894.

Aschauer, D. F., Kreuz, S., & Rumpel, S. (2013). Analysis of transduction efficiency, tropism and axonal transport of AAV serotypes 1, 2, 5, 6, 8 and 9 in the mouse brain. *PloS one*, *8*(9), e76310.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., & Harris, M. A. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, *25*(1), 25.

Astic, L., Saucier, D., Coulon, P., Lafay, F., & Flamand, A. (1993). The CVS strain of rabies virus as transneuronal tracer in the olfactory system of mice. *Brain research*, *619*(1-2), 146-156.

Atchison, R. W., Casto, B. C., & Hammon, W. M. (1965). Adenovirus-associated defective virus particles. *Science*, *149*(3685), 754-755.

Auer, H., Lyianarachchi, S., Newsom, D., Klisovic, M. I., & Kornacker, K. (2003). Chipping away at the chip bias: RNA degradation in microarray analysis. *Nature genetics*, *35*(4), 292.

Bachevalier, J., & Nemanic, S. (2008). Memory for spatial location and object-place associations are differently processed by the hippocampal formation, parahippocampal areas TH/TF and perirhinal cortex. *Hippocampus*, *18*(1), 64-80.

Bailey, C. H., Bartsch, D., & Kandel, E. R. (1996). Toward a molecular definition of long-term memory storage. *Proceedings of the National Academy of Sciences*, *93*(24), 13445-13452.

Bains, J. S., & Oliet, S. H. (2007). Glia: they make your memories stick!. *Trends in neurosciences*, *30*(8), 417-424.

Balderas, I., Rodriguez-Ortiz, C. J., Salgado-Tonda, P., Chavez-Hurtado, J., McGaugh, J. L., & Bermudez-Rattoni, F. (2008). The consolidation of object and context recognition memory involve different regions of the temporal lobe.*Learning & Memory*, *15*(9), 618-624.

Bansal, V. (2017). A computational method for estimating the PCR duplication rate in DNA and RNA-seq experiments. *BMC bioinformatics*, *18*(3), 43.

Barker, G. R., Banks, P. J., Scott, H., Ralph, G. S., Mitrophanous, K. A., Wong, L. F., Bashir, Z.I., Uney, J.B., & Warburton, E. C. (2017). Separate elements of episodic memory subserved by distinct hippocampal–prefrontal connections. *Nature neuroscience*, *20*(2), 242.

Barker, G. R., Bird, F., Alexander, V., & Warburton, E. C. (2007). Recognition memory for objects, place, and temporal order: a disconnection analysis of the role of the medial prefrontal cortex and perirhinal cortex. *Journal of Neuroscience*, *27*(11), 2948-2957.

Barker, G. R., Warburton, E. C., Koder, T., Dolman, N. P., More, J. C., Aggleton, J. P., Bashir, Z.I., Auberson, Y.P., Jane, D.E., & Brown, M. W. (2006). The different effects on recognition memory of perirhinal kainate and NMDA glutamate receptor antagonism: implications for underlying plasticity mechanisms. *The Journal of neuroscience*, *26*(13), 3561-3566.

Barker, G. R., & Warburton, E. C. (2008). NMDA receptor plasticity in the perirhinal and prefrontal cortices is crucial for the acquisition of long-term object-in-place associative memory. *The Journal of neuroscience*, *28*(11), 2837-2844.

Barker, G. R., & Warburton, E. C. (2011). When is the hippocampus involved in recognition memory?. *The Journal of Neuroscience*, *31*(29), 10721-10731.

Barnes, P., Kirtley, A., & Thomas, K. L. (2012). Quantitatively and qualitatively different cellular processes are engaged in CA1 during the consolidation and reconsolidation of contextual fear memory. *Hippocampus*, *22*(2), 149-171.

Barondes, S. H., & Cohen, H. D. (1967). Delayed and sustained effect of acetoxycycloheximide on memory in mice. *Proc Natl Acad Sci USA, 58*, 157–164.

Barr, D. S., Lambert, N. A., Hoyt, K. L., Moore, S. D., & Wilson, W. A. (1995). Induction and reversal of long-term potentiation by low-and high-intensity theta pattern stimulation. *Journal of Neuroscience*, *15*(7), 5402-5410.

Barria, A., Muller, D., Derkach, V., Griffith, L. C., & Soderling, T. R. (1997). Regulatory phosphorylation of AMPA-type glutamate receptors by CaM-KII during long-term potentiation. *Science*, *276*(5321), 2042-2045.

Barry, D. N., Coogan, A. N., & Commins, S. (2016). The time course of systems consolidation of spatial memory from recent to remote retention: A comparison of the Immediate Early Genes Zif268, c-Fos and Arc. *Neurobiology of learning and memory*, *128*, 46-55.

Bartel, D. P., Sheng, M., Lau, L. F., & Greenberg, M. E. (1989). Growth factors and membrane depolarization activate distinct programs of early response gene expression: dissociation of fos and jun induction. *Genes & development*, *3*(3), 304-313.

Bartholomae, C. C., Arens, A., Balaggan, K. S., Yáñez-Muñoz, R. J., Montini, E., Howe, S. J., Paruzynski, A., Korn, B., Appelt, J.U., MacNeil, A., & Cesana, D. (2011). Lentiviral vector integration profiles differ in rodent postmitotic tissues. *Molecular Therapy*, *19*(4), 703-710.

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise?. *Frontiers in psychology*, *4*, 907.

Baruzzo, G., Hayer, K. E., Kim, E. J., Di Camillo, B., FitzGerald, G. A., & Grant, G. R. (2017). Simulation-based comprehensive benchmarking of RNA-seq aligners. *nAture methods*, *14*(2), 135.

Bashir, Z. I., Bortolotto, Z. A., Davies, C. H., Berretta, N., Irving, A. J., Seal, A. J., Henley, J.M., Jane, D.E., Watkins, J.C., & Collingridge, G. L. (1993). Induction of LTP in the hippocampus needs synaptic activation of glutamate metabotropic receptors. *Nature*, *363*(6427), 347.

Bazan, N. G. (2003). Synaptic lipid signaling significance of polyunsaturated fatty acids and platelet-activating factor. *Journal of lipid research*, *44*(12), 2221-2233.

Beattie, E. C., Carroll, R. C., Yu, X., Morishita, W., Yasuda, H., von Zastrow, M., & Malenka, R. C. (2000). Regulation of AMPA receptor endocytosis by a signaling mechanism shared with LTD. *Nature neuroscience*, *3*(12), 1291.

Beier, K. T., Saunders, A., Oldenburg, I. A., Miyamichi, K., Akhtar, N., Luo, L., Whelan, S.P., Sabatini, B., & Cepko, C. L. (2011). Anterograde or retrograde transsynaptic labeling of CNS neurons with vesicular stomatitis virus vectors. *Proceedings of the National Academy of Sciences*, *108*(37), 15414-15419.

Bekinschtein, P., Cammarota, M., Igaz, L. M., Bevilaqua, L. R., Izquierdo, I., & Medina, J. H. (2007). Persistence of long-term memory storage requires a late protein synthesis-and BDNF-dependent phase in the hippocampus. *Neuron*,*53*(2), 261-277.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Benson, D. L., Schnapp, L. M., Shapiro, L., & Huntley, G. W. (2000). Making memories stick: cell-adhesion molecules in synaptic plasticity. *Trends in cell biology*, *10*(11), 473-482.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., & Boutell, J. M. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *nature*, *456*(7218), 53.

Berrettini, W. H., Vuoristo, J., Ferraro, T. N., Buono, R. J., Wildenauer, D., & Ala-Kokko, L. (1998). Human G (olf) gene polymorphisms and vulnerability to bipolar disorder. *Psychiatric genetics*, *8*(4), 235-238.

Bertaina-Anglade, V., Tramu, G., & Destrade, C. (2000). Differential learning-stage dependent patterns of c-Fos protein expression in brain regions during the acquisition and memory

consolidation of an operant task in mice. *European Journal of Neuroscience*, *12*(10), 3803-3812.

Bezaire, M. J., & Soltesz, I. (2013). Quantitative assessment of CA1 local circuits: knowledge base for interneuron-pyramidal cell connectivity. *Hippocampus*, *23*(9), 751-785.

Bi, G. Q., & Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of neuroscience*, *18*(24), 10464-10472.

Billia, F., Baskys, A., Carlen, P. L., & De Boni, U. (1992). Rearrangement of centromeric satellite DNA in hippocampal neurons exhibiting long-term potentiation. *Molecular brain research*, *14*(1-2), 101-108.

Bird, C. M., & Burgess, N. (2008). The hippocampus and memory: insights from spatial processing. *Nature Reviews Neuroscience*, *9*(3), nrn2335.

Bito, H., Deisseroth, K., & Tsien, R. W. (1996). CREB phosphorylation and dephosphorylation: a Ca2+-and stimulus duration–dependent switch for hippocampal gene expression. *Cell*, *87*(7), 1203-1214.

Blainey, P., Krzywinski, M., & Altman, N. (2014). Points of significance: replication. *Nature methods*, *11*(9), 879.

Blessing, D., & Déglon, N. (2016). Adeno-associated virus and lentivirus vectors: a refined toolkit for the central nervous system. *Current opinion in virology*, *21*, 61-66.

Bliss, T. V., & Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature*, *361*(6407), 31.

Bliss, T. V., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of physiology*, *232*(2), 331-356.

Blitzer, R. D., Connor, J. H., Brown, G. P., Wong, T., Shenolikar, S., Iyengar, R., & Landau, E. M. (1998). Gating of CaMKii by cAMP-regulated protein phosphatase activity during LTP. *Science*, *280*(5371), 1940-1943.

Blömer, U., Naldini, L., Kafri, T., Trono, D., Verma, I. M., & Gage, F. H. (1997). Highly efficient and sustained gene transfer in adult neurons with a lentivirus vector. *Journal of virology*, *71*(9), 6641-6649.

Blum, K. I., & Abbott, L. F. (1996). A model of spatial map formation in the hippocampus of the rat. *Neural computation*, *8*(1), 85-93.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.

Bolognani, F., & Perrone-Bizzozero, N. I. (2008). RNA–protein interactions and control of mRNA stability in neurons. *Journal of neuroscience research*, *86*(3), 481-489.

Bondy, S. C. (1966). The ribonucleic acid metabolism of the brain. *Journal of neurochemistry*, *13*(10), 955-959.

Bosch, M., & Hayashi, Y. (2012). Structural plasticity of dendritic spines. *Current opinion in neurobiology*, *22*(3), 383-388.

Bouard, D., Alazard-Dany, N., & Cosset, F. L. (2009). Viral vectors: from virology to transgene expression. *British journal of pharmacology*, *157*(2), 153-165.

Bourtchuladze, R., Abel, T., Berman, N., Gordon, R., Lapidus, K., & Kandel, E. R. (1998). Different training procedures recruit either one or two critical periods for contextual memory consolidation, each of which requires protein synthesis and PKA. *Learning & Memory*, *5*(4), 365-374.

Bradlow, E. T. (2002). Exploring repeated measures data sets for key features using Principal Components Analysis. *International Journal of Research in Marketing*, *19*(2), 167-179.

Bradshaw, K. D., Emptage, N. J., & Bliss, T. V. P. (2003). A role for dendritic protein synthesis in hippocampal late LTP. *European Journal of Neuroscience*, *18*(11), 3150-3152.

Bramham, C. R., & Wells, D. G. (2007). Dendritic mRNA: transport, translation and function. *Nature Reviews Neuroscience*, *8*(10), 776.

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, *25*(2), 163-177.

Bredt, D. S., & Nicoll, R. A. (2003). AMPA receptor trafficking at excitatory synapses. *Neuron*, *40*(2), 361-379.

Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: what are the roles of the perirhinal cortex and hippocampus?. *Nature Reviews Neuroscience*,*2*(1), 51-61.

Brown, M. W., Barker, G. R. I., Aggleton, J. P., & Warburton, E. C. (2012). What pharmacological interventions indicate concerning the role of the perirhinal cortex in recognition memory. *Neuropsychologia*, *50*(13), 3122-3140.

Bullard, J. H., Purdom, E., Hansen, K. D., & Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC bioinformatics*, *11*(1), 94.

Bunting, K. M., Nalloor, R. I., & Vazdarjanova, A. (2016). Influence of Isoflurane on Immediate-Early Gene Expression. *Frontiers in behavioral neuroscience*, *9*, 363.

Burger, C., Gorbatyuk, O. S., Velardo, M. J., Peden, C. S., Williams, P., Zolotukhin, S., Reier, P.J., Mandel, R.J., & Muzyczka, N. (2004). Recombinant AAV viral vectors pseudotyped with viral capsids from serotypes 1, 2, and 5 display differential efficiency and cell tropism after delivery to different regions of the central nervous system. *Molecular Therapy*, *10*(2), 302-317.

Burrone, J., O'byrne, M., & Murthy, V. N. (2002). Multiple forms of synaptic plasticity triggered by selective suppression of activity in individual neurons. *Nature*, *420*(6914), 414.

Burwell, R. D., & Amaral, D. G. (1998). Cortical afferents of the perirhinal, postrhinal, and entorhinal cortices of the rat. *Journal of Comparative Neurology*, *398*(2), 179-205.

Bussey, T. J., Duck, J., Muir, J. L., & Aggleton, J. P. (2000). Distinct patterns of behavioural impairments resulting from fornix transection or neurotoxic lesions of the perirhinal and postrhinal cortices in the rat. *Behavioural brain research*, *111*(1-2), 187-202.

Cahill, L. (1997). The neurobiology of emotionally influenced memory implications for understanding traumatic memory. *Annals of the New York Academy of Sciences*, *821*(1), 238-246.

Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., Landthaler, M., Obermayer, B., & Ohler, U. (2016). Detecting actively translated open reading frames in ribosome profiling data. *Nature methods*, *13*(2), 165.

Camon, E., Magrane, M., Barrell, D., Lee, V., Dimmer, E., Maslen, J., Binns, D., Harte, N., Lopez, R., & Apweiler, R. (2004). The gene ontology annotation (goa) database: sharing knowledge in uniprot with gene ontology. *Nucleic acids research*, *32*(suppl_1), D262-D266.

Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L. A., & Johnson, R. (2016). Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *Rna*, *22*(6), 867-882.

Caroni, P., Donato, F., & Muller, D. (2012). Structural plasticity upon learning: regulation and functions. *Nature Reviews Neuroscience*, *13*(7), nrn3258.

Carter, B. J. (2004). Adeno-associated virus and the development of adeno-associated virus vectors: a historical perspective. *Molecular therapy*, *10*(6), 981-989.

Castle, M. J., Gershenson, Z. T., Giles, A. R., Holzbaur, E. L., & Wolfe, J. H. (2014). Adeno-associated virus serotypes 1, 8, and 9 share conserved mechanisms for anterograde and retrograde axonal transport. *Human gene therapy*, *25*(8), 705-720.

Catterall, W. A., & Few, A. P. (2008). Calcium channel regulation and presynaptic plasticity. *Neuron*, *59*(6), 882-901.

Cavallaro, S., D'Agata, V., Manickam, P., Dufour, F., & Alkon, D. L. (2002). Memory-specific temporal profiles of gene expression in the hippocampus. *Proceedings of the National Academy of Sciences*, *99*(25), 16279-16284.

Cearley, C. N., Vandenberghe, L. H., Parente, M. K., Carnish, E. R., Wilson, J. M., & Wolfe, J. H. (2008). Expanded repertoire of AAV vector serotypes mediate unique patterns of transduction in mouse brain. *Molecular therapy*, *16*(10), 1710-1718.

Cearley, C. N., & Wolfe, J. H. (2006). Transduction characteristics of adeno-associated virus vectors expressing cap serotypes 7, 8, 9, and Rh10 in the mouse brain. *Molecular Therapy*, *13*(3), 528-537.

Chen, H., & Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC bioinformatics*, *12*(1), 35.

Chen, P. B., Kawaguchi, R., Blum, C., Achiro, J. M., Coppola, G., O'Dell, T. J., & Martin, K. C. (2017). Mapping gene expression in excitatory neurons during hippocampal late-phase long-term potentiation. *Frontiers in molecular neuroscience*, *10*, 39.

Chen, J., Park, C. S., & Tang, S. J. (2006). Activity-dependent synaptic Wnt release regulates hippocampal long term potentiation. *Journal of Biological Chemistry*, *281*(17), 11910-11916.

Cheval, H., Chagneau, C., Levasseur, G., Veyrac, A., Faucon-Biguet, N., Laroche, S., & Davis, S. (2012). Distinctive features of Egr transcription factor regulation and DNA binding activity in CA1 of the hippocampus in synaptic plasticity and consolidation and reconsolidation of fear memory. *Hippocampus*, *22*(3), 631-642.

Cho, E., Kim, D. H., Hur, Y. N., Whitcomb, D. J., Regan, P., Hong, J. H., Kim, H., Suh, Y.H., Cho, K., & Park, M. (2015a). Cyclin Y inhibits plasticity-induced AMPA receptor exocytosis and LTP. *Scientific reports*, *5*, 12624.

Cho, J., Yu, N. K., Choi, J. H., Sim, S. E., Kang, S. J., Kwak, C., Lee, S.W., Kim, J.I., Choi, D.I., Kim, V.N., & Kaang, B. K. (2015b). Multiple repressive mechanisms in the hippocampus during memory formation. *Science*, *350*(6256), 82-87.

Chowdhury, S., Shepherd, J. D., Okuno, H., Lyford, G., Petralia, R. S., Plath, N., Kuhl, D., Huganir, R.L., & Worley, P. F. (2006). Arc/Arg3. 1 interacts with the endocytic machinery to regulate AMPA receptor trafficking. *Neuron*, *52*(3), 445-459.

Citri, A., & Malenka, R. C. (2008). Synaptic plasticity: multiple forms, functions, and mechanisms. *Neuropsychopharmacology*, *33*(1), 18.

Clark, R. E., Zola, S. M., & Squire, L. R. (2000). Impaired recognition memory in rats after damage to the hippocampus. *Journal of Neuroscience*, *20*(23), 8853-8860.

Cmarko, D., Smigova, J., Minichova, L., & Popov, A. (2008). Nucleolus, The ribosome factory. *Histology & histopathology, 23*(10), 1291-1298.

Coba, M. P., Valor, L. M., Kopanitsa, M. V., Afinowi, N. O., & Grant, S. G. (2008). Kinase networks integrate profiles of N-methyl-D-aspartate receptor-mediated gene expression in hippocampus. *Journal of Biological Chemistry*, *283*(49), 34101-34107.

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, Amnesia, and the Hippocampal System.* Cambridge: MIT Press.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, *210*(4466), 207-210.

Cohen, S. J., & Stackman Jr, R. W. (2015). Assessing rodent hippocampal involvement in the novel object recognition task. A review. *Behavioural brain research*, *285*, 105-117.

Cole, A. J., Saffen, D. W., Baraban, J. M., & Worley, P. F. (1989). Rapid increase of an immediate early gene messenger RNA in hippocampal neurons by synaptic NMDA receptor activation. *Nature*, *340*(6233), 474.

Colvin, R. A., Fontaine, C. P., Laskowski, M., & Thomas, D. (2003). Zn2+ transporters and Zn2+ homeostasis in neurons. *European journal of pharmacology*, *479*(1-3), 171-185.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome biology*, *17*(1), 13.

Cooke, S. F., & Bear, M. F. (2015). Visual recognition memory: a view from V1. *Current opinion in neurobiology*, *35*, 57-65.

Cook-Snyder, D. R., Jones, A., & Reijmers, L. G. (2015). A retrograde adeno-associated virus for collecting ribosome-bound mRNA from anatomically defined projection neurons. *Frontiers in molecular neuroscience*, *8*, 56.

Coso, O. A., Chiariello, M., Yu, J. C., Teramoto, H., Crespo, P., Xu, N., Miki, T., & Gutkind, J. S. (1995). The small GTP-binding proteins Rac1 and Cdc42regulate the activity of the JNK/SAPK signaling pathway. *Cell*, *81*(7), 1137-1146.

Costa-Mattioli, M., Gobert, D., Stern, E., Gamache, K., Colina, R., Cuello, C., Sossin, W., Kaufman, R., Pelletier, J., Rosenblum, K., & Krnjević, K. (2007). eIF2α phosphorylation bidirectionally regulates the switch from short-to long-term synaptic plasticity and memory. *Cell*, *129*(1), 195-206.

Costa-Mattioli, M., Sossin, W. S., Klann, E., & Sonenberg, N. (2009). Translational control of long-lasting synaptic plasticity and memory. *Neuron*, *61*(1), 10-26.

Crosio, C., Heitz, E., Allis, C. D., Borrelli, E., & Sassone-Corsi, P. (2003). Chromatin remodeling and neuronal response: multiple signaling pathways induce specific histone H3 modifications and early gene expression in hippocampal neurons. *Journal of Cell Science*, *116*(24), 4905-4914.

Cross, L., Brown, M. W., Aggleton, J. P., & Warburton, E. C. (2013). The medial dorsal thalamic nucleus and the medial prefrontal cortex of the rat function together to support associative recognition and recency but not item recognition. *Learning & memory*, *20*(1), 41-50.

Dash, P. K., Hochner, B., & Kandel, E. R. (1990). Injection of the cAMP-responsive element into the nucleus of Aplysia sensory neurons blocks long-term facilitation. *Nature*, *345*(6277), 718.

Davis, G. W., & Goodman, C. S. (1998). Genetic analysis of synaptic development and plasticity: homeostatic regulation of synaptic efficacy. *Current opinion in neurobiology*, *8*(1), 149-156.

Davis, H. P., & Squire, L. R. (1984). Protein synthesis and memory: a review. *Psychological bulletin*, *96*(3), 518.

Davis, E. K., Zou, Y., & Ghosh, A. (2008). Wnts acting through canonical and noncanonical signaling pathways exert opposite effects on hippocampal synapse formation. *Neural development*, *3*(1), 32.

Dayton, R. D., Wang, D. B., Cain, C. D., Schrott, L. M., Ramirez, J. J., King, M. A., & Klein, R. L. (2012b). Frontotemporal lobar degeneration-related proteins induce only subtle memory-related deficits when bilaterally overexpressed in the dorsal hippocampus. *Experimental neurology*, *233*(2), 807-814.

Dayton, R. D., Wang, D. B., & Klein, R. L. (2012a). The advent of AAV9 expands applications for brain and spinal cord gene delivery. *Expert opinion on biological therapy*, *12*(6), 757-766.

Debiec, J., LeDoux, J. E., & Nader, K. (2002). Cellular and systems reconsolidation in the hippocampus. *Neuron*, *36*(3), 527-538.

Deng, W., Aimone, J. B., & Gage, F. H. (2010). New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory? *Nature reviews neuroscience*, *11*(5), 339.

Deniz, E., & Erman, B. (2017). Long noncoding RNA (lincRNA), a new paradigm in gene expression control. *Functional & integrative genomics*, *17*(2-3), 135-143.

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., & Lempicki, R. A. (2003). DAVID: database for annotation, visualization, and integrated discovery. Genome biology, 4(9), 1.

Dent, E. W., Merriam, E. B., & Hu, X. (2011). The dynamic cytoskeleton: backbone of dendritic spine plasticity. *Current opinion in neurobiology*, *21*(1), 175-181.

Desmaris, N., Bosch, A., Salaün, C., Petit, C., Prévost, M. C., Tordo, N., Perrin, P., Schwartz, O., De Rocquigny, H., & Heard, J. M. (2001). Production and neurotropism of lentivirus vectors pseudotyped with lyssavirus envelope glycoproteins. *Molecular Therapy*, *4*(2), 149-156.

Devaraju, P., Yu, J., Eddins, D., Mellado-Lagarde, M. M., Earls, L. R., Westmoreland, J. J., Quarato, G., Green, D.R., & Zakharenko, S. S. (2017). Haploinsufficiency of the 22q11. 2 microdeletion gene Mrpl40 disrupts short-term synaptic plasticity and working memory through dysregulation of mitochondrial calcium. *Molecular psychiatry*, *22*(9), 1313.

Dever, T. E. (2002). Gene-specific regulation by general translation factors. *Cell*, *108*(4), 545-556.

D'haeseleer, P., Liang, S., & Somogyi, R. (2000). Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics*, *16*(8), 707-726.

Di Giovanni, S., Molon, A., Broccolini, A., Melcon, G., Mirabella, M., Hoffman, E. P., & Servidei, S. (2004). Constitutive activation of MAPK cascade in acute quadriplegic myopathy. *Annals of neurology*, *55*(2), 195-206.

Dijkmans, T. F., van Hooijdonk, L. W. A., Schouten, T. G., Kamphorst, J. T., Fitzsimons, C. P., & Vreugdenhil, E. (2009). Identification of new Nerve Growth Factor-responsive immediate-early genes. *Brain research*, *1249*, 19-33.

Dillies, M. A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., & Guernec, G. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in bioinformatics*, *14*(6), 671-683.

Dittgen, T., Nimmerjahn, A., Komai, S., Licznerski, P., Waters, J., Margrie, T. W., Helmchen, F., Denk, W., Brecht, M., & Osten, P. (2004). Lentivirus-based genetic manipulations of cortical neurons and their optical and electrophysiological monitoring in vivo.*Proceedings of the National Academy of Sciences*, *101*(52), 18206-18211.

Dix, S. L., & Aggleton, J. P. (1999). Extending the spontaneous preference test of recognition: evidence of object-location and object-context recognition. *Behavioural brain research*, *99*(2), 191-200.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21.

Donahue, C. P., Jensen, R. V., Ochiishi, T., Eisenstein, I., Zhao, M., Shors, T., & Kosik, K. S. (2002). Transcriptional profiling reveals regulated genes in the hippocampus during memory formation. *Hippocampus*, *12*(6), 821-833.

Donato, F., Rompani, S. B., & Caroni, P. (2013). Parvalbumin-expressing basket-cell network plasticity induced by experience regulates adult learning. *Nature*, *504*(7479), 272.

Donello, J. E., Loeb, J. E., & Hope, T. J. (1998). Woodchuck hepatitis virus contains a tripartite posttranscriptional regulatory element. *Journal of virology*, *72*(6), 5085-5092.

Dong, J. Y., Fan, P. D., & Frizzell, R. A. (1996). Quantitative analysis of the packaging capacity of recombinant adeno-associated virus. *Human gene therapy*, *7*(17), 2101-2112.

Dong, H. W., Swanson, L. W., Chen, L., Fanselow, M. S., & Toga, A. W. (2009). Genomic–anatomic evidence for distinct functional domains in hippocampal field CA1. *Proceedings of the National Academy of Sciences*, *106*(28), 11794-11799.

Doudna, J. A., & Sarnow, P. (2007). Translation initiation by viral internal ribosome entry sites. *Cold Spring Harbor Monograph Series*, *48*, 129.

Doyle, J. P., Dougherty, J. D., Heiman, M., Schmidt, E. F., Stevens, T. R., Ma, G., Bupp, S., Shrestha, P., Shah, R.D., Doughty, M.L., & Heintz, N. (2008). Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell*, *135*(4), 749-762.

Drane, L., Ainsley, J. A., Mayford, M. R., & Reijmers, L. G. (2014). A transgenic mouse line for collecting ribosome-bound mRNA using the tetracycline transactivator system. *Frontiers in molecular neuroscience*, *7*.

Dudai, Y., & Morris, R. G. (2000). To consolidate or not to consolidate: what are the questions. *Brain, perception, memory. Advances in cognitive sciences*, 149-162.

Dudek, S. M., & Bear, M. F. (1992). Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(10), 4363.

Dull, T., Zufferey, R., Kelly, M., Mandel, R. J., Nguyen, M., Trono, D., & Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. *Journal of virology*, *72*(11), 8463-8471.

Eacker, S. M., Crawford, K., Brichta, L., Riessland, M., Ingolia, N. T., Greengard, P., ... & Dawson, V. L. (2017). Experience-dependent translational state defined by cell type-specific ribosome profiling. *bioRxiv*, 169425.

Earnshaw, W. C., Sullivan, K. F., Machlin, P. S., Cooke, C. A., Kaiser, D. A., Pollard, T. D., ... & Cleveland, D. W. (1987). Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. *The Journal of cell biology*, *104*(4), 817-829.

Edelmann, E., Cepeda-Prado, E., & Leßmann, V. (2017). Coexistence of multiple types of synaptic plasticity in individual hippocampal CA1 pyramidal neurons. *Frontiers in synaptic neuroscience*, *9*, 7.

Eichenbaum, H. (2014). Time cells in the hippocampus: a new dimension for mapping memories. *Nature Reviews Neuroscience*, *15*(11), 732.

Eichenbaum, H., & Cohen, N. J. (2014). Can we reconcile the declarative memory and spatial navigation views on hippocampal function?. *Neuron*, *83*(4), 764-770.

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, *95*(25), 14863-14868.

Ekstrand, M. I., Nectow, A. R., Knight, Z. A., Latcha, K. N., Pomeranz, L. E., & Friedman, J. M. (2014). Molecular profiling of neurons based on connectivity.*Cell*, *157*(5), 1230-1242.

Engström, P. G., Steijger, T., Sipos, B., Grant, G. R., Kahles, A., Alioto, T., ... & Davis, C. A. (2013). Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature methods*, *10*(12), 1185.

Ennaceur, A., & Delacour, J. (1988). A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data. *Behavioural brain research*, *31*(1), 47-59.

Erlwein, O., Robinson, M. J., Dustan, S., Weber, J., Kaye, S., & McClure, M. O. (2011). DNA extraction columns contaminated with murine sequences. *PLoS One*, *6*(8), e23484.

Fanselow, M. S., & Dong, H. W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures?. *Neuron*, *65*(1), 7-19.

Fernández, E., Collins, M. O., Uren, R. T., Kopanitsa, M. V., Komiyama, N. H., Croning, M. D., ... & Grant, S. G. (2009). Targeted tandem affinity purification of PSD-95 recovers core

postsynaptic complexes and schizophrenia susceptibility proteins. *Molecular systems biology*, *5*(1), 269.

Fernandez, E. M., Díaz-Ceso, M. D., & Vilar, M. (2015). Brain expressed and X-linked (Bex) proteins are intrinsically disordered proteins (IDPs) and form new signaling hubs. *PloS one*, *10*(1), e0117

Ferrer, I., Garcia-Esparcia, P., Carmona, M., Carro, E., Aronica, E., Kovacs, G. G., ... & Gustincich, S. (2016). Olfactory receptors in non-chemosensory organs: the nervous system in health and disease. *Frontiers in aging neuroscience*, *8*, 163.

Filipowicz, W., Bhattacharyya, S. N., & Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?. *Nature reviews genetics*, *9*(2), 102.

Fioravante, D., & Byrne, J. H. (2011). Protein degradation and memory formation. *Brain research bulletin*, *85*(1-2), 14-20.

Fischer, A., Sananbenesi, F., Schrick, C., Spiess, J., & Radulovic, J. (2002). Cyclin-dependent kinase 5 is required for associative learning. *Journal of Neuroscience*, *22*(9), 3700-3707.

Flavell, S. W., & Greenberg, M. E. (2008). Signaling mechanisms linking neuronal activity to gene expression and plasticity of the nervous system. *Annu. Rev. Neurosci.*, *31*, 563-590.

Fleischmann, A., Hvalby, O., Jensen, V., Strekalova, T., Zacher, C., Layer, L. E., ... & Gass, P. (2003). Impaired long-term memory and NR2A-type NMDA receptor-dependent synaptic plasticity in mice lacking c-Fos in the CNS. *The journal of neuroscience*, *23*(27), 9116-9122.

Fonseca, R., Vabulas, R. M., Hartl, F. U., Bonhoeffer, T., & Nägerl, U. V. (2006). A balance of protein synthesis and proteasome-dependent degradation determines the maintenance of LTP. *Neuron*, *52*(2), 239-245.

Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature neuroscience*, *5*(5), 458.

Fortin, N. J., Wright, S. P., & Eichenbaum, H. (2004). Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature*, *431*(7005), 188.

Frank, C. L., & Tsai, L. H. (2009). Alternative functions of core cell cycle regulators in neuronal migration, neuronal maturation, and synaptic plasticity. *Neuron*, *62*(3), 312-326.

Freund, T. F., & Buzsáki, G. (1996). Interneurons of the hippocampus. *Hippocampus*, *6*(4), 347-470.

Frey, U., Huang, Y. Y., & Kandel, E. R. (1993). Effects of cAMP simulate a late stage of LTP in hippocampal CA1 neurons. *Science*, *260*(5114), 1661-1664.

Frey, U., & Morris, R. G. (1997). Synaptic tagging and long-term potentiation. *Nature*, *385*(6616), 533-536.

Fujita, Y., & Yamashita, T. (2017). Spatial organization of genome architecture in neuronal development and disease. *Neurochemistry international*.

Fusi, S., Drew, P. J., & Abbott, L. F. (2005). Cascade models of synaptically stored memories. *Neuron*, *45*(4), 599-611.

Fyhn, M., Molden, S., Hollup, S., Moser, M. B., & Moser, E. I. (2002). Hippocampal neurons responding to first-time dislocation of a target object. *Neuron*, *35*(3), 555-566.

Gaba, A., Wang, Z., Krishnamoorthy, T., Hinnebusch, A. G., & Sachs, M. S. (2001). Physical evidence for distinct mechanisms of translational control by upstream open reading frames. *The EMBO journal*, *20*(22), 6453-6463.

Gagnon-Bartsch, J. A., & Speed, T. P. (2012). Using control genes to correct for unwanted variation in microarray data. *Biostatistics*, *13*(3), 539-552.

Gazzaniga, M. S. (1966). Interhemispheric communication of visual learning. *Neuropsychologia*, *4*(2), 183-189.

Gene Ontology Consortium. (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic acids research*, *45*(D1), D331-D338.

Gentleman, V. J., Carey, D. M., Bates, B., Bolstad, M., Dettling, S., Dudoit, B., Ellis, L., Gautier, Y. Ge, and others. (2004). Bioconductor: Open software development for computational biology and bioinformatics R. *Genome Biology, Vol. 5,* R80.

Gerashchenko, M. V., & Gladyshev, V. N. (2014). Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic acids research*, *42*(17), e134-e134.

Ghirardi, M., Montarolo, P. G., & Kandel, E. R. (1995). A novel intermediate stage in the transition between short-and long-term facilitation in the sensory to motor neuron synapse of Aplysia. *Neuron*, *14*(2), 413-420.

Giese, K. P., Aziz, W., Kraev, I., & Stewart, M. G. (2015). Generation of multi-innervated dendritic spines as a novel mechanism of long-term memory formation. *Neurobiology of learning and memory*, *124*, 48-51.

Giese, K. P., Fedorov, N. B., Filipkowski, R. K., & Silva, A. J. (1998). Autophosphorylation at Thr286 of the α calcium-calmodulin kinase II in LTP and learning. *Science*, *279*(5352), 870-873.

Gingras, A. C., Gygi, S. P., Raught, B., Polakiewicz, R. D., Abraham, R. T., Hoekstra, M. F., ... & Sonenberg, N. (1999). Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes & development*, *13*(11), 1422-1437.

Goeman, J. J., & Bühlmann, P. (2007). Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics*, *23*(8), 980-987.

Gold, P. E., & McGaugh, J. L. (1975). A single-trace, two-process view of memory storage processes. *Short-term memory*, 355-378.

Goldie, B. J., & Cairns, M. J. (2012). Post-transcriptional trafficking and regulation of neuronal gene expression. *Molecular neurobiology*, *45*(1), 99-108.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, *15*(1), 20-25.

Gothard, K. M., Skaggs, W. E., & McNaughton, B. L. (1996). Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues. *Journal of Neuroscience*, *16*(24), 8027-8040.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Chen, Z. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, *29*(7), 644.

Gräff, J., Koshibu, K., Jouvenceau, A., Dutar, P., & Mansuy, I. M. (2010). Protein phosphatase 1-dependent transcriptional programs for long-term memory and plasticity. *Learning & memory*, *17*(7), 355-363.

Gray, S. J., Foti, S. B., Schwartz, J. W., Bachaboina, L., Taylor-Blake, B., Coleman, J., ... & Samulski, R. J. (2011). Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors. *Human gene therapy*, *22*(9), 1143-1153.

Grecksch, G., & Matthies, H. (1980). Two sensitive periods for the amnesic effect of anisomycin. *Pharmacology Biochemistry and Behavior*, *12*(5), 663-665.

Greengard, P., Allen, P. B., & Nairn, A. C. (1999). Beyond the Dopamine Receptor: the DARPP-32/Protein Phosphatase-1 Cascade. *Neuron*, *23*(3), 435-447.

Gross, S. R., & Kinzy, T. G. (2005). Translation elongation factor 1A is essential for regulation of the actin cytoskeleton and cell morphology. *Nature Structural and Molecular Biology*, *12*(9), 772.

Grossmann, S., Bauer, S., Robinson, P. N., & Vingron, M. (2007). Improved detection of overrepresentation of Gene-Ontology annotations with parent–child analysis. *Bioinformatics*, *23*(22), 3024-3031.

Gu, J., Firestein, B. L., & Zheng, J. Q. (2008). Microtubules in dendritic spine development. *Journal of Neuroscience*, *28*(46), 12120-12124.

Guan, Z., Giustetto, M., Lomvardas, S., Kim, J. H., Miniaci, M. C., Schwartz, J. H., ... & Kandel, E. R. (2002). Integration of long-term-memory-related synaptic plasticity involves bidirectional regulation of gene expression and chromatin structure. *Cell*, *111*(4), 483-493.

Gundersen, G. W., Jagodnik, K. M., Woodland, H., Fernandez, N. F., Sani, K., Dohlman, A. B., ... & Ma'ayan, A. (2016). GEN3VA: aggregation and analysis of gene expression signatures from related studies. *BMC bioinformatics*, *17*(1), 461.

Guzowski, J. F. (2002). Insights into immediate-early gene function in hippocampal memory consolidation using antisense oligonucleotide and fluorescent imaging approaches. *Hippocampus*, *12*(1), 86-104.

Guzowski, J. F., Lyford, G. L., Stevenson, G. D., Houston, F. P., McGaugh, J. L., Worley, P. F., & Barnes, C. A. (2000). Inhibition of activity-dependent arc protein expression in the rat hippocampus impairs the maintenance of long-term potentiation and the consolidation of long-term memory. *The Journal of neuroscience*, *20*(11), 3993-4001.

Guzowski, J. F., McNaughton, B. L., Barnes, C. A., & Worley, P. F. (1999). Environment-specific expression of the immediate-early gene Arc in hippocampal neuronal ensembles. *Nature neuroscience*, *2*(12), 1120.

Guzowski, J. F., Setlow, B., Wagner, E. K., & McGaugh, J. L. (2001). Experience-dependent gene expression in the rat hippocampus after spatial learning: a comparison of the immediate-early genesArc, c-fos, and zif268. *The Journal of Neuroscience*, *21*(14), 5089-5098.

Haditsch, U., Leone, D. P., Farinelli, M., Chrostek-Grashoff, A., Brakebusch, C., Mansuy, I. M., ... & Palmer, T. D. (2009). A central role for the small GTPase Rac1 in hippocampal plasticity and spatial learning and memory. *Molecular and Cellular Neuroscience*, *41*(4), 409-419.

Haettig, J., Sun, Y., Wood, M. A., & Xu, X. (2013). Cell-type specific inactivation of hippocampal CA1 disrupts location-dependent object recognition in the mouse. *Learning & memory*, *20*(3), 139-146.

Hall, A. (1998). Rho GTPases and the actin cytoskeleton. *Science*, *279*(5350), 509-514.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in cognitive sciences*, *5*(9), 394-400.

Hammer, R. E., Maika, S. D., Richardson, J. A., Tang, J. P., & Taurog, J. D. (1990). Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human β2m: an animal model of HLA-B27-associated human disorders. *Cell*, *63*(5), 1099-1112.

Hamra, F. K., Gatlin, J., Chapman, K. M., Grellhesl, D. M., Garcia, J. V., Hammer, R. E., & Garbers, D. L. (2002). Production of transgenic rats by lentiviral transduction of male germ-line stem cells. *Proceedings of the national academy of sciences*, *99*(23), 14931-14936.

Han, H. J., Allen, C. C., Buchovecky, C. M., Yetman, M. J., Born, H. A., Marin, M. A., ... & Jankowsky, J. L. (2012). Strain background influences neurotoxicity and behavioral abnormalities in mice expressing the tetracycline transactivator. *The Journal of Neuroscience*, *32*(31), 10574-10586.

Han, K., Kim, M. H., Seeburg, D., Seo, J., Verpelli, C., Han, S., ... & Do Heo, W. (2009). Regulated RalBP1 binding to RalA and PSD-95 controls AMPA receptor endocytosis and LTD. *PLoS biology*, *7*(9), e1000187.

Hansen, K. D., Brenner, S. E., & Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic acids research*, *38*(12), e131-e131.

Hardcastle, T. J., & Kelly, K. A. (2010). baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC bioinformatics*, *11*(1), 422.

Hardcastle, T. J., & Kelly, K. A. (2013). Empirical Bayesian analysis of paired high-throughput sequencing data with a beta-binomial distribution. *BMC bioinformatics*, *14*(1), 135.

Hardingham, G. E., Arnold, F. J., & Bading, H. (2001). Nuclear calcium signaling controls CREB-mediated gene expression triggered by synaptic activity. *Nature neuroscience*, *4*(3), 261.

Hardy, S., Kitamura, M., Harris-Stansil, T., Dai, Y., & Phipps, M. L. (1997). Construction of adenovirus vectors through Cre-lox recombination. *Journal of virology*, *71*(3), 1842-1849.

Harris, R. S. (2007). *Improved pairwise alignment of genomic DNA*. PhD thesis; Pennsylvania State University.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100-108.

Harvey, C. D., & Svoboda, K. (2007). Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature*, *450*(7173), 1195.

Hashimoto, T. B., Edwards, M. D., & Gifford, D. K. (2014). Universal count correction for high-throughput sequencing. *PLoS computational biology*, *10*(3), e1003494.

Hawasli, A. H., Benavides, D. R., Nguyen, C., Kansy, J. W., Hayashi, K., Chambon, P., Greengard, P., Powell, C.M., Cooper, D.C., & Bibb, J. A. (2007). Cyclin-dependent kinase 5 governs learning and synaptic plasticity via control of NMDAR degradation. *Nature neuroscience*, *10*(7), 880.

Hayashi, Y., Shi, S. H., Esteban, J. A., Piccini, A., Poncer, J. C., & Malinow, R. (2000). Driving AMPA receptors into synapses by LTP and CaMKii: requirement for GluR1 and PDZ domain interaction. *Science*, *287*(5461), 2262-2267.

Hayer, K. E., Pizarro, A., Lahens, N. F., Hogenesch, J. B., & Grant, G. R. (2015). Benchmark analysis of algorithms for determining and quantifying full-length mRNA splice forms from RNA-seq data. *Bioinformatics*, *31*(24), 3938-3945.

He, X., & Zhang, J. (2006). Why do hubs tend to be essential in protein networks?. *PLoS genetics*, *2*(6), e88.

Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: overviews and challenges. *Biotechniques*, *56*(2), 61.

Hedrick, N. G., Harward, S. C., Hall, C. E., Murakoshi, H., McNamara, J. O., & Yasuda, R. (2016). Rho GTPase complementation underlies BDNF-dependent homo-and heterosynaptic plasticity. *Nature*, *538*(7623), 104.

Heer, J., Card, S. K., & Landay, J. A. (2005, April). Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 421-430). ACM.

Heiman, M., Schaefer, A., Gong, S., Peterson, J. D., Day, M., Ramsey, K. E., ... & Heintz, N. (2008). A translational profiling approach for the molecular characterization of CNS cell types. *Cell*, *135*(4), 738-748.

Heiman, M., Kulicke, R., Fenster, R. J., Greengard, P., & Heintz, N. (2014). Cell type–specific mRNA purification by translating ribosome affinity purification (TRAP). *Nature protocols*, *9*(6), 1282-1291.

Heintz, N. (2001). BAC to the future: the use of bac transgenic mice for neuroscience research. *Nature Reviews Neuroscience*, *2*(12), 861.

Hermens, W. T., & Verhaagen, J. (1998). Viral vectors, tools for gene transfer in the nervous system. *Progress in neurobiology*, *55*(4), 399-432.

Hermonat, P. L., & Muzyczka, N. (1984). Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells. *Proceedings of the National Academy of Sciences*, *81*(20), 6466-6470.

Heroux, N. A., Osborne, B. F., Miller, L. A., Kawan, M., Buban, K. N., Rosen, J. B., & Stanton, M. E. (2018). Differential expression of the immediate early genes c-Fos, Arc, Egr-1, and Npas4 during long-term memory formation in the context preexposure facilitation effect (CPFE). *Neurobiology of learning and memory*, *147*, 128-138.

Herzog, E., Takamori, S., Jahn, R., Brose, N., & Wojcik, S. M. (2006). Synaptic and vesicular co-localization of the glutamate transporters VGLUT1 and VGLUT2 in the mouse hippocampus. *Journal of neurochemistry*, *99*(3), 1011-1018.

Hioki, H., Kameda, H., Nakamura, H., Okunomiya, T., Ohira, K., Nakamura, K., ... & Kaneko, T. (2007). Efficient gene transduction of neurons by lentivirus with enhanced neuron-specific promoters. *Gene therapy*, *14*(11), 872.

Ho, D. Y., & Mocarski, E. S. (1988). β-Galactosidase as a marker in the peripheral and neural tissues of the herpes simplex virus-infected mouse. *Virology*, *167*(1), 279-283.

Hoge, J., & Kesner, R. P. (2007). Role of CA3 and CA1 subregions of the dorsal hippocampus on temporal processing of objects. *Neurobiology of learning and memory*, *88*(2), 225-231.

Holehonnur, R., Luong, J. A., Chaturvedi, D., Ho, A., Lella, S. K., Hosek, M. P., & Ploski, J. E. (2014). Adeno-associated viral serotypes produce differing titers and differentially transduce neurons within the rat basal and lateral amygdala. *BMC neuroscience*, *15*(1), 28.

Hölscher, C., Anwyl, R., & Rowan, M. J. (1997). Stimulation on the positive phase of hippocampal theta rhythm induces long-term potentiation that can be depotentiated by stimulation on the negative phase in area CA1 in vivo. *Journal of Neuroscience*, *17*(16), 6470-6477.

Hong, S. J., Li, H., Becker, K. G., Dawson, V. L., & Dawson, T. M. (2004). Identification and analysis of plasticity-induced late-response genes. *Proceedings of the National Academy of Sciences*, *101*(7), 2145-2150.

Hooijmans, C. R., & Kiliaan, A. J. (2008). Fatty acids, lipid metabolism and Alzheimer pathology. *European journal of pharmacology*, *585*(1), 176-196.

Horne, M. R., Iordanova, M. D., Albasser, M. M., Aggleton, J. P., Honey, R. C., & Pearce, J. M. (2010). Lesions of the perirhinal cortex do not impair integration of visual and geometric information in rats. *Behavioral neuroscience*, *124*(3), 311.

Howard, D. B., Powers, K., Wang, Y., & Harvey, B. K. (2008). Tropism and toxicity of adeno-associated viral vector serotypes 1, 2, 5, 6, 7, 8, and 9 in rat neurons and glia in vitro. *Virology*, *372*(1), 24-34.

Howarth, J. L., Lee, Y. B., & Uney, J. B. (2010). Using viral vectors as gene transfer tools (Cell Biology and Toxicology Special Issue: ETCS-UK 1 day meeting on genetic manipulation of cells). *Cell biology and toxicology*, *26*(1), 1-20.

Hu, X., Gadbury, G. L., Xiang, Q., & Allison, D. B. (2010a). Illustrations on Using the Distribution of a P-value in High Dimensional Data Analyses. *Advances and applications in statistical sciences*, *1*(2), 191.

Hu, J. H., Park, J. M., Park, S., Xiao, B., Dehoff, M. H., Kim, S., ... & Linden, D. J. (2010b). Homeostatic scaling requires group I mGluR activation mediated by Homer1a. *Neuron*, *68*(6), 1128-1142.

Huang, Y., Ainsley, J. A., Reijmers, L. G., & Jackson, F. R. (2013). Translational profiling of clock cells reveals circadianly synchronized protein synthesis.

Huang, D. W., Sherman, B. T., Tan, Q., Kir, J., Liu, D., Bryant, D., ... & Lempicki, R. A. (2007). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic acids research*, *35*(suppl_2), W169-W175.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*(1), 106-154.

Huber, K. M., Kayser, M. S., & Bear, M. F. (2000). Role for rapid dendritic protein synthesis in hippocampal mGluR-dependent long-term depression. *Science*, *288*(5469), 1254-1256.

Hunsaker, M. R., & Kesner, R. P. (2008). Evaluating the differential roles of the dorsal dentate gyrus, dorsal CA3, and dorsal CA1 during a temporal ordering for spatial locations task. *Hippocampus*, *18*(9), 955-964.

Iacono, G., Altafini, C., & Torre, V. (2013). Early phase of plasticity-related gene regulation and SRF dependent transcription in the hippocampus. *PloS one*, *8*(7), e68078.

Ibata, K., Sun, Q., & Turrigiano, G. G. (2008). Rapid synaptic scaling induced by changes in postsynaptic firing. *Neuron*, *57*(6), 819-826.

Igaz, L. M., Vianna, M. R., Medina, J. H., & Izquierdo, I. (2002). Two time periods of hippocampal mRNA synthesis are required for memory consolidation of fear-motivated learning. *The Journal of neuroscience*, *22*(15), 6781-6789.

Igaz, L. M., Bekinschtein, P., Vianna, M. M., Izquierdo, I., & Medina, J. H. (2004). Gene expression during memory formation. *Neurotoxicity research*, *6*(3), 189-203.

Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., Harris, M. S., Talhouarne, G. J., Jackson, S. E., ... & Weissman, J. S. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell reports*, *8*(5), 1365-1379.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *science*, *324*(5924), 218-223.

Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, *147*(4), 789-802.

Izquierdo, I., Bevilaqua, L. R., Rossato, J. I., Bonini, J. S., Medina, J. H., & Cammarota, M. (2006). Different molecular cascades in different sites of the brain control memory consolidation. *Trends in neurosciences*, *29*(9), 496-505.

Izquierdo, I., & Medina, J. H. (1997). Memory formation: the sequence of biochemical events in the hippocampus and its connection to activity in other brain structures. *Neurobiology of learning and memory*, *68*(3), 285-316.

Jain, P., & Bhalla, U. S. (2014). Transcription control pathways decode patterned synaptic inputs into diverse mRNA expression profiles. *PloS one*,*9*(5).

Jakobsson, J., Ericson, C., Jansson, M., Björk, E., & Lundberg, C. (2003). Targeted transgene expression in rat brain using lentiviral vectors. *Journal of neuroscience research*, *73*(6), 876-885.

Jamal, M., Ameno, K., Miki, T., Tanaka, N., Ono, J., Shirakami, G., ... & Kinoshita, H. (2012). High ethanol and acetaldehyde impair spatial memory in mouse models: opposite effects of aldehyde dehydrogenase 2 and apolipoprotein E on memory. *Pharmacology Biochemistry and Behavior*, *101*(3), 443-449.

Jaworski, J., Kapitein, L. C., Gouveia, S. M., Dortland, B. R., Wulf, P. S., Grigoriev, I., ... & Krugers, H. (2009). Dynamic microtubules regulate dendritic spine morphology and synaptic plasticity. *Neuron*, *61*(1), 85-100.

Jeewajee, A., Lever, C., Burton, S., O'keefe, J., & Burgess, N. (2008). Environmental novelty is signaled by reduction of the hippocampal theta frequency. *Hippocampus*, *18*(4), 340-348.

Jeffery, G. (1984). Retinal ganglion cell death and terminal field retraction in the developing rodent visual system. *Developmental Brain Research*, *13*(1), 81-96.

Jiang, D., Tang, C., & Zhang, A. (2004). Cluster analysis for gene expression data: A survey. *IEEE Transactions on knowledge and data engineering*, *16*(11), 1370-1386.

Johannes, G., Carter, M. S., Eisen, M. B., Brown, P. O., & Sarnow, P. (1999). Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F

concentrations using a cDNA microarray. *Proceedings of the National Academy of Sciences*, *96*(23), 13118-13123.

Joo, J. Y., Schaukowitch, K., Farbiak, L., Kilaru, G., & Kim, T. K. (2016). Stimulus-specific combinatorial functionality of neuronal c-fos enhancers. *Nature neuroscience*, *19*(1), 75.

Josselyn, S. A., Köhler, S., & Frankland, P. W. (2015). Finding the engram. *Nature Reviews Neuroscience*, *16*(9), 521.

Jouvenceau, A., Hédou, G., Potier, B., Kollen, M., Dutar, P., & Mansuy, I. M. (2006). Partial inhibition of PP1 alters bidirectional synaptic plasticity in the hippocampus. *European Journal of Neuroscience*, *24*(2), 564-572.

Ju, W., Morishita, W., Tsui, J., Gaietta, G., Deerinck, T. J., Adams, S. R., ... & Malenka, R. C. (2004). Activity-dependent regulation of dendritic synthesis and trafficking of AMPA receptors. *Nature neuroscience*, *7*(3), 244.

Jung, I., Jo, K., Kang, H., Ahn, H., Yu, Y., & Kim, S. (2017). TimesVector: a vectorized clustering approach to the analysis of time series transcriptome data from multiple phenotypes. *Bioinformatics*, *33*(23), 3827-3835.

Jung, J., & Jung, H. (2016). Methods to analyze cell type-specific gene expression profiles from heterogeneous cell populations. *Animal Cells and Systems*, *20*(3), 113-117.

Jung, M. W., Wiener, S. I., & McNaughton, B. L. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *Journal of Neuroscience*, *14*(12), 7347-7356.

Kandel, E. R., Dudai, Y., & Mayford, M. R. (2014). The molecular and systems biology of memory. *Cell*, *157*(1), 163-186.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27-30.

Kaneko, M., Stellwagen, D., Malenka, R. C., & Stryker, M. P. (2008). Tumor necrosis factor-α mediates one component of competitive, experience-dependent plasticity in developing visual cortex. *Neuron*, *58*(5), 673-680.

Kang, H., & Schuman, E. M. (1996). A requirement for local protein synthesis in neurotrophin-induced hippocampal synaptic plasticity. *Science*, *273*(5280), 1402-1406.

Kapitein, L. C., Schlager, M. A., Kuijpers, M., Wulf, P. S., van Spronsen, M., MacKintosh, F. C., & Hoogenraad, C. C. (2010). Mixed microtubules steer dynein-driven cargo transport into dendrites. *Current Biology*, *20*(4), 290-299.

Kaplitt, M. G., Leone, P., Samulski, R. J., Xiao, X., Pfaff, D. W., O'Malley, K. L., & During, M. J. (1994). Long-term gene expression and phenotypic correction using adeno-associated virus vectors in the mammalian brain. *Nature genetics*, *8*(2), 148.

Kasai, H., Matsuzaki, M., Noguchi, J., Yasumatsu, N., & Nakahara, H. (2003). Structure–stability–function relationships of dendritic spines. *Trends in neurosciences*, *26*(7), 360-368.

Katche, C., Bekinschtein, P., Slipczuk, L., Goldin, A., Izquierdo, I. A., Cammarota, M., & Medina, J. H. (2010). Delayed wave of c-Fos expression in the dorsal hippocampus involved specifically in persistence of long-term memory storage. *Proceedings of the National Academy of Sciences*, *107*(1), 349-354.

Kato, S., Kuramochi, M., Takasumi, K., Kobayashi, K., Inoue, K. I., Takahara, D., ... & Kobayashi, K. (2011). Neuron-specific gene transfer through retrograde transport of lentiviral vector pseudotyped with a novel type of fusion envelope glycoprotein. *Human gene therapy*, *22*(12), 1511-1523.

Keck, T., Toyoizumi, T., Chen, L., Doiron, B., Feldman, D. E., Fox, K., ... & Lisman, J. E. (2017). Integrating Hebbian and homeostatic plasticity: the current state of the field and future research directions. *Phil. Trans. R. Soc. B*, *372*(1715), 20160158.

Kemp, A., & Manahan-Vaughan, D. (2004). Hippocampal long-term depression and long-term potentiation encode different aspects of novelty acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(21), 8192-8197.

Kemp, A., & Manahan-Vaughan, D. (2007). Hippocampal long-term depression: master or minion in declarative memory processes?. *Trends in neurosciences*, *30*(3), 111-118.

Kemp, A., & Manahan-Vaughan, D. (2012). Passive spatial perception facilitates the expression of persistent hippocampal long-term depression. *Cerebral cortex*, *22*(7), 1614-1621.

Keyser, D. O., & Pellmar, T. C. (1994). Synaptic transmission in the hippocampus: critical role for glial cells. *Glia*, *10*(4), 237-243.

Kim, J., Delcasso, S., & Lee, I. (2011). Neural correlates of object-in-place learning in hippocampus and prefrontal cortex. *Journal of Neuroscience*, *31*(47), 16991-17006.

Kim, J. J., Fanselow, M. S., DeCola, J. P., & Landeira-Fernandez, J. (1992). Selective impairment of long-term but not short-term conditional fear by the N-methyl-D-aspartate antagonist APV. *Behavioral neuroscience*, *106*(4), 591.

Kim, T. K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., ... & Markenscoff-Papadimitriou, E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, *465*(7295), 182.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, *14*(4), R36.

Kim, I. H., Wang, H., Soderling, S. H., & Yasuda, R. (2014). Loss of Cdc42 leads to defects in synaptic plasticity and remote memory recall. *Elife*, *3*.

King, H. A., & Gerber, A. P. (2014). Translatome profiling: methods for genome-scale analysis of mRNA translation. *Briefings in functional genomics*, *15*(1), 22-31.

Kinnavane, L., Albasser, M. M., & Aggleton, J. P. (2015). Advances in the behavioural testing and network imaging of rodent recognition memory. *Behavioural brain research*, *285*, 67-78.

Klausberger, T., & Somogyi, P. (2008). Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science*, *321*(5885), 53-57.

Klein, R. L., Dayton, R. D., Leidenheimer, N. J., Jansen, K., Golde, T. E., & Zweig, R. M. (2006). Efficient neuronal gene transfer with AAV8 leads to neurotoxic levels of tau or green fluorescent proteins. *Molecular Therapy*, *13*(3), 517-527.

Klein, R. L., Dayton, R. D., Tatom, J. B., Henderson, K. M., & Henning, P. P. (2008). AAV8, 9, Rh10, Rh43 vector gene transfer in the rat brain: effects of serotype, promoter and purification method. *Molecular therapy*, *16*(1), 89-96.

Klein, R. L., Hamby, M. E., Gong, Y., Hirko, A. C., Wang, S., Hughes, J. A., ... & Meyer, E. M. (2002). Dose and promoter effects of adeno-associated viral vector for green fluorescent protein expression in the rat brain. *Experimental neurology*, *176*(1), 66-74.

Klein, R. L., Meyer, E. M., Peel, A. L., Zolotukhin, S., Meyers, C., Muzyczka, N., & King, M. A. (1998). Neuron-specific transduction in the rat septohippocampal or nigrostriatal pathway by recombinant adeno-associated virus vectors. *Experimental neurology*, *150*(2), 183-194.

Klepikova, A. V., Kasianov, A. S., Chesnokov, M. S., Lazarevich, N. L., Penin, A. A., & Logacheva, M. (2017). Effect of method of deduplication on estimation of differential gene expression using RNA-seq. *PeerJ, 5*, e3091.

Knierim, J. J., Lee, I., & Hargreaves, E. L. (2006). Hippocampal place cells: parallel input streams, subregional processing, and implications for episodic memory. *Hippocampus*, *16*(9), 755-764.

Knight, Z. A., Tan, K., Birsoy, K., Schmidt, S., Garrison, J. L., Wysocki, R. W., ... & Friedman, J. M. (2012). Molecular profiling of activated neurons by phosphorylated ribosome capture. *Cell*, *151*(5), 1126-1137.

Kohara, K., Pignatelli, M., Rivest, A. J., Jung, H. Y., Kitamura, T., Suh, J., ... & Wickersham, I. R. (2014). Cell type–specific genetic and optogenetic tools reveal hippocampal CA2 circuits. *Nature neuroscience*, *17*(2), 269.

Komorowski, R. W., Manns, J. R., & Eichenbaum, H. (2009). Robust conjunctive item–place coding by hippocampal neurons parallels learning what happens where. *Journal of Neuroscience*, *29*(31), 9918-9929.

Korzus, E., Rosenfeld, M. G., & Mayford, M. (2004). CBP histone acetyltransferase activity is a critical component of memory consolidation. *Neuron*, *42*(6), 961-972.

Kotaleski, J. H., & Blackwell, K. T. (2010). Modelling the molecular mechanisms of synaptic plasticity using systems biology approaches. *Nature Reviews Neuroscience*, *11*(4), nrn2807.

Koudinov, A. R., & Koudinova, N. V. (2001). Essential role for cholesterol in synaptic plasticity and neuronal degeneration. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, *15*(10), 1858.

Kratz, A., Beguin, P., Kaneko, M., Chimura, T., Suzuki, A. M., Matsunaga, A., ... & Launey, T. (2014). Digital expression profiling of the compartmentalized translatome of Purkinje neurons. *Genome research*, *24*(8), 1396-1410.

Kügler, S., Lingor, P., Schöll, U., Zolotukhin, S., & Bähr, M. (2003). Differential transgene expression in brain cells in vivo and in vitro from AAV-2 vectors with small transcriptional control units. *Virology*, *311*(1), 89-95.

Kumaran, D., & Maguire, E. A. (2006). An unexpected sequence of events: mismatch detection in the human hippocampus. *PLoS biology*, *4*(12), e424.

Kuroda, H., Kutner, R. H., Bazan, N. G., & Reiser, J. (2008). A comparative analysis of constitutive and cell-specific promoters in the adult mouse hippocampus using lentivirus vector-mediated gene transfer. *The journal of gene medicine*, *10*(11), 1163-1175.

Lanahan, A., & Worley, P. (1998). Immediate-early genes and synaptic function. *Neurobiology of learning and memory*, *70*(1), 37-43.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357.

Langmore, J. P. (2002). Rubicon Genomics, Inc. *Pharmacogenomics*, *3*(4), 557-560.

Langston, R. F., & Wood, E. R. (2010). Associative recognition and the hippocampus: Differential effects of hippocampal lesions on object-place, object-context and object-place-context memory. *Hippocampus*, *20*(10), 1139-1153.

Larkin, M. C., Lykken, C., Tye, L. D., Wickelgren, J. G., & Frank, L. M. (2014). Hippocampal output area CA1 broadcasts a generalized novelty signal during an object-place recognition task. *Hippocampus*, *24*(7), 773-783.

Laurin, N., Ickowicz, A., Pathare, T., Malone, M., Tannock, R., Schachar, R., ... & Barr, C. L. (2008). Investigation of the G protein subunit Gαolf gene (GNAL) in attention deficit/hyperactivity disorder. *Journal of psychiatric research*, *42*(2), 117-124.

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology*, *15*(2), R29.

Leach, P. T., Poplawski, S. G., Kenney, J. W., Hoffman, B., Liebermann, D. A., Abel, T., & Gould, T. J. (2012). Gadd45b knockout mice exhibit selective deficits in hippocampus-dependent long-term memory. *Learning & Memory*, *19*(8), 319-324.

Lee, H. K., Barbarosie, M., Kameyama, K., Bear, M. F., & Huganir, R. L. (2000). Regulation of distinct AMPA receptor phosphorylation sites during bidirectional synaptic plasticity. *Nature*, *405*(6789), 955.

Lee, I., Hunsaker, M. R., & Kesner, R. P. (2005). The role of hippocampal subregions in detecting spatial novelty. *Behavioral neuroscience*, *119*(1), 145.

Lee, H. K., Kameyama, K., Huganir, R. L., & Bear, M. F. (1998). NMDA induces long-term synaptic depression and dephosphorylation of the GluR1 subunit of AMPA receptors in hippocampus. *Neuron*, *21*(5), 1151-1162.

Lee, I., & Park, S. B. (2013). Perirhinal cortical inactivation impairs object-in-place memory and disrupts task-dependent firing in hippocampal CA1, but not in CA3. *Frontiers in neural circuits*, *7*, 134.

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., ... & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, *11*(10), 733.

Leek, J. T., & Storey, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS genetics*, *3*(9), e161.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., & Chen, L. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, *445*(7124), 168.

Lein, E. S., Zhao, X., & Gage, F. H. (2004). Defining a molecular atlas of the hippocampus using DNA microarrays and high-throughput in situ hybridization. The Journal of neuroscience, 24(15), 3879-3889.

Lenck-Santini, P. P., Rivard, B., Muller, R. U., & Poucet, B. (2005). Study of CA1 place cell activity and exploratory behavior following spatial and nonspatial changes in the environment. *Hippocampus*, *15*(3), 356-369.

Levenson, J. M., O'Riordan, K. J., Brown, K. D., Trinh, M. A., Molfese, D. L., & Sweatt, J. D. (2004). Regulation of histone acetylation during memory formation in the hippocampus. *Journal of Biological Chemistry*, *279*(39), 40545-40559.

Lever, C., Burton, S., Jeewajee, A., Wills, T. J., Cacucci, F., Burgess, N., & O'keefe, J. (2010). Environmental novelty elicits a later theta phase of firing in CA1 but not subiculum. *Hippocampus*, *20*(2), 229-234.

Lewis, B. P., Green, R. E., & Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences*, *100*(1), 189-192.

Li, J., & Bushel, P. R. (2016). EPIG-Seq: extracting patterns and identifying co-expressed genes from RNA-Seq data. *BMC genomics*, *17*(1), 255.

Li, B., Carey, M., & Workman, J. L. (2007). The role of chromatin during transcription. *Cell*, *128*(4), 707-719.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Li, S., Łabaj, P. P., Zumbo, P., Sykacek, P., Shi, W., Shi, L., ... & Thierry-Mieg, D. (2014). Detecting and correcting systematic variation in large-scale RNA sequencing data. *Nature biotechnology*, *32*(9), 888.

Li, Z., Okamoto, K. I., Hayashi, Y., & Sheng, M. (2004). The importance of dendritic mitochondria in the morphogenesis and plasticity of spines and synapses. *Cell*, *119*(6), 873-887.

Li, J., & Tibshirani, R. (2013). Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Statistical methods in medical research*, *22*(5), 519-536.

Lin, Y., Bloodgood, B. L., Hauser, J. L., Lapan, A. D., Koon, A. C., Kim, T. K., ... & Greenberg, M. E. (2008). Activity-dependent regulation of inhibitory synapse development by Npas4. *Nature*, *455*(7217), 1198.

Lin, Y., Golovnina, K., Chen, Z. X., Lee, H. N., Negron, Y. L. S., Sultana, H., ... & Harbison, S. T. (2016). Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual Drosophila melanogaster. *BMC genomics*, *17*(1), 28.

Lisman, J. (2017). Glutamatergic synapses are structurally and biochemically complex because of multiple plasticity processes: long-term potentiation, long-term depression, short-term potentiation and scaling. *Phil. Trans. R. Soc. B*, *372*(1715), 20160260.

Lisman, J. E., & Otmakhova, N. A. (2001). Storage, recall, and novelty detection of sequences by the hippocampus: elaborating on the SOCRATIC model to account for normal and aberrant effects of dopamine. *Hippocampus*, *11*(5), 551-568.

Lisman, J., Schulman, H., & Cline, H. (2002). The molecular basis of CaMKii function in synaptic and behavioural memory. *Nature Reviews Neuroscience*, *3*(3), 175.

Lisman, J. E., & Zhabotinsky, A. M. (2001). A model of synaptic memory: a CaMKii/PP1 switch that potentiates transmission by organizing an AMPA receptor anchoring assembly. *Neuron*, *31*(2), 191-201.

Liu, Y., Du, S., Lv, L., Lei, B., Shi, W., Tang, Y., ... & Zhong, Y. (2016). Hippocampal activation of Rac1 regulates the forgetting of object recognition memory. *Current Biology*, *26*(17), 2351-2357.

Liu, R., Lei, J. X., Luo, C., Lan, X., Chi, L., Deng, P., ... & Liu, Q. Y. (2012a). Increased EID1 nuclear translocation impairs synaptic plasticity and memory function associated with pathogenesis of Alzheimer's disease. *Neurobiology of disease*, *45*(3), 902-912.

Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K., & Tonegawa, S. (2012b). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*, *484*(7394), 381.

Liu, P., & Si, Y. (2014). Cluster analysis of RNA-sequencing data. In *Statistical Analysis of Next Generation Sequencing Data* (pp. 191-217). Springer International Publishing.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). Viruses: Structure, Function, and Uses. (In Molecular cell biology, 4th ed).

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*(12), 550.

Ludin, B., & Matus, A. (1993). The neuronal cytoskeleton and its role in axonal and dendritic plasticity. *Hippocampus*, *3*(S1), 61-71.

Lukas, T. J., Mirzoeva, S., & Watterson, D. M. (1998). Calmodulin-regulated protein kinases. *Calmodulin and signal transduction*, 65-168.

Lun, A. T., Chen, Y., & Smyth, G. K. (2016). It's DE-licious: a recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR. *Statistical Genomics: Methods and Protocols*, 391-416.

Lüscher, C., Xia, H., Beattie, E. C., Carroll, R. C., von Zastrow, M., Malenka, R. C., & Nicoll, R. A. (1999). Role of AMPA receptor cycling in synaptic transmission and plasticity. *Neuron*, *24*(3), 649-658.

Lynch, G., Kramár, E. A., & Gall, C. M. (2015). Protein synthesis and consolidation of memory-related synaptic changes. *Brain research, 1621.* 62-72.

Lynch, G., Larson, J., Kelso, S., Barrionuevo, G., & Schottler, F. (1983). Intracellular injections of EGTA block induction of hippocampal long-term potentiation. *Nature*, *305*(5936), 719.

Lyons, M. R., & West, A. E. (2011). Mechanisms of specificity in neuronal activity-regulated gene transcription. *Progress in neurobiology*, *94*(3), 259-295.

Ma, L., Bayram, Y., McLaughlin, H. M., Cho, M. T., Krokosky, A., Turner, C. E., ... & Bodurtha, J. (2016). De novo missense variants in PPP1CB are associated with intellectual disability and congenital heart disease. *Human genetics*, *135*(12), 1399-1409.

Ma, D. K., Jang, M. H., Guo, J. U., Kitabatake, Y., Chang, M. L., Pow-Anpongkul, N., ... & Song, H. (2009). Neuronal activity–induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science*, *323*(5917), 1074-1077.

Ma, H., Li, B., & Tsien, R. W. (2015). Distinct roles of multiple isoforms of CaMKii in signaling to the nucleus. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1853*(9), 1953-1957.

Ma'ayan, A., Jenkins, S. L., Neves, S., Hasseldine, A., Grace, E., Dubin-Thaler, B., ... & Kershenbaum, A. (2005). Formation of regulatory patterns during signal propagation in a mammalian cellular network. *Science*, *309*(5737), 1078-1083.

MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, *71*(4), 737-749.

MacDonald, J. F., Jackson, M. F., & Beazely, M. A. (2006). Hippocampal long-term synaptic plasticity and signal amplification of NMDA receptors. *Critical Reviews™ in Neurobiology*, *18*(1-2).

Madabhushi, R., Gao, F., Pfenning, A. R., Pan, L., Yamakawa, S., Seo, J., ... & Stott, R. T. (2015). Activity-induced DNA breaks govern the expression of neuronal early-response genes. *Cell*, *161*(7), 1592-1605.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2017). cluster: Cluster Analysis Basics and Extensions. R package version 2.0. 1. 2015.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, *97*(8), 4398-4403.

Maguschak, K. A., & Ressler, K. J. (2012). The dynamic role of beta-catenin in synaptic plasticity. *Neuropharmacology*, *62*(1), 78-88.

Mahajan, G., & Mande, S. C. (2015). From system-wide differential gene expression to perturbed regulatory factors: a combinatorial approach. *PloS one*, *10*(11), e0142147.

Malenka, R. C., & Bear, M. F. (2004). LTP and LTD: an embarrassment of riches. *Neuron*, *44*(1), 5-21.

Malenka, R. C., Kauer, J. A., Zucker, R. S., & Nicoll, R. A. (1988). Postsynaptic calcium is sufficient for potentiation of hippocampal synaptic transmission. *Science*, *242*(4875), 81-84.

Malik, A. N., Vierbuchen, T., Hemberg, M., Rubin, A. A., Ling, E., Couch, C. H., ... & Greenberg, M. E. (2014). Genome-wide identification and characterization of functional neuronal activity–dependent enhancers. *Nature neuroscience*, *17*(10), 1330.

Manahan-Vaughan, D., & Braunewell, K. H. (1999). Novelty acquisition is associated with induction of hippocampal long-term depression. *Proceedings of the National Academy of Sciences*, *96*(15), 8739-8744.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological review*, *87*(3), 252.

Marder, E., & Prinz, A. A. (2002). Modeling stability in neuron and network function: the role of activity in homeostasis. *Bioessays*, *24*(12), 1145-1154.

Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, *275*(5297), 213-215.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., & Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research*, *18*(9), 1509-1517.

Marr, D. (1971). Simple Memory: A Theory for Archicortex. *Philosophical Transactions of the Royal Society of London Series B*, *262*, 23-81.

Martin, K. C., & Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. *Cell*, *136*(4), 719-730.

Martin, S. J., Grimwood, P. D., & Morris, R. G. M. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis. *Annual review of neuroscience*, *23*(1), 649-711.

Martin, K. C., Michael, D., Rose, J. C., Barad, M., Casadio, A., Zhu, H., & Kandel, E. R. (1997). MAP kinase translocates into the nucleus of the presynaptic cell and is required for long-term facilitation in Aplysia. *Neuron*, *18*(6), 899-912.

Masters, B. A., Quaife, C. J., Erickson, J. C., Kelly, E. J., Froelick, G. J., Zambrowicz, B. P., ... & Palmiter, R. D. (1994). Metallothionein III is expressed in neurons that sequester zinc in synaptic vesicles. *Journal of Neuroscience*, *14*(10), 5844-5857.

Matsuzaki, M., Honkura, N., Ellis-Davies, G. C., & Kasai, H. (2004). Structural basis of long-term potentiation in single dendritic spines. *Nature*, *429*(6993), 761.

Matthies, H., Frey, U., Reymann, K., Krug, M., Jork, R., & Schroeder, H. (1990). Different mechanisms and multiple stages of LTP. In *Excitatory Amino Acids and Neuronal Plasticity* (pp. 359-368). Springer, Boston, MA.

Mattson, M. P., Gleichmann, M., & Cheng, A. (2008). Mitochondria in neuroplasticity and neurological disorders. *Neuron*, *60*(5), 748-766.

Mayes, A. R., Holdstock, J. S., Isaac, C. L., Hunkin, N. M., & Roberts, N. (2002). Relative sparing of item recognition memory in a patient with adult-onset damage limited to the hippocampus. *Hippocampus*, *12*(3), 325-340.

Mayford, M., Wang, J., Kandel, E. R., & O'Dell, T. J. (1995). CaMKii regulates the frequency-response function of hippocampal synapses for the production of both LTD and LTP. *Cell*, *81*(6), 891-904.

Mazarakis, N. D., Azzouz, M., Rohll, J. B., Ellard, F. M., Wilkes, F. J., Olsen, A. L., ... & Kingsman, A. J. (2001). Rabies virus glycoprotein pseudotyping of lentiviral vectors enables retrograde axonal transport and access to the nervous system after peripheral delivery. *Human molecular genetics*, *10*(19), 2109-2121.

Maze, I., Wenderski, W., Noh, K. M., Bagot, R. C., Tzavaras, N., Purushothaman, I., ... & Tamminga, C. A. (2015). Critical role of histone turnover in neuronal transcription and plasticity. *Neuron*, *87*(1), 77-94.

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research*, *40*(10), 4288-4297.

McCarty, D. M., Young Jr, S. M., & Samulski, R. J. (2004). Integration of adeno-associated virus (AAV) and recombinant AAV vectors. *Annu. Rev. Genet.*, *38*, 819-845.

McFarland, N. R., Lee, J. S., Hyman, B. T., & McLean, P. J. (2009). Comparison of transduction efficiency of recombinant AAV serotypes 1, 2, 5, and 8 in the rat nigrostriatal system. *Journal of neurochemistry*, *109*(3), 838-845.

McIntyre, C. K., Miyashita, T., Setlow, B., Marjon, K. D., Steward, O., Guzowski, J. F., & McGaugh, J. L. (2005). Memory-influencing intra-basolateral amygdala drug infusions modulate expression of Arc protein in the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(30), 10718-10723.

McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., & Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, *83*(1), 202-215.

McLean, J. R., Smith, G. A., Rocha, E. M., Hayes, M. A., Beagan, J. A., Hallett, P. J., & Isacson, O. (2014). Widespread neuron-specific transgene expression in brain and spinal cord

following synapsin promoter-driven AAV9 neonatal intracerebroventricular injection. *Neuroscience letters*, *576*, 73-78.

McNaughton, B. L., & Morris, R. G. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in neurosciences*, *10*(10), 408-415.

Meijer, H. U., & Thomas, A. (2002). Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA.*Biochem. J*, *367*, 1-11.

Mello, C., Nottebohm, F., & Clayton, D. (1995). Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene's response to that song in zebra finch telencephalon. *Journal of Neuroscience*, *15*(10), 6919-6925.

Mellor, J., Nicoll, R. A., & Schmitz, D. (2002). Mediation of hippocampal mossy fiber long-term potentiation by presynaptic Ih channels. *Science*, *295*(5552), 143-147.

Mendez, M., Arias, N., Uceda, S., & Arias, J. L. (2015). c-Fos expression correlates with performance on novel object and novel place recognition tests. *Brain research bulletin*, *117*, 16-23.

Merico, D., Isserlin, R., Stueker, O., Emili, A., & Bader, G. D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PloS one*, *5*(11), e13984.

Meyuhas, O., & Dreazen, A. (2009). Ribosomal protein S6 kinase: from TOP mRNAs to cell size. *Progress in molecular biology and translational science*, *90*, 109-153.

Meyuhas, O., & Kahan, T. (2015). The race to decipher the top secrets of TOP mRNAs. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1849*(7), 801-811.

Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., Lewis, S., & Thomas, P. D. (2010). PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic acids research*, *38*(suppl_1), D204-D210.

Michod, D., Bartesaghi, S., Khelifi, A., Bellodi, C., Berliocchi, L., Nicotera, P., & Salomoni, P. (2012). Calcium-dependent dephosphorylation of the histone chaperone DAXX regulates H3. 3 loading and transcription upon neuronal activation. *Neuron*, *74*(1), 122-135.

Miklos, G. L. G., & Rubin, G. M. (1996). The role of the genome project in determining gene function: insights from model organisms. *Cell*, *86*(4), 521-529.

Milekic, M. H., & Alberini, C. M. (2002). Temporally graded requirement for protein synthesis following memory reactivation. *Neuron*, *36*(3), 521-525.

Milivojevic, B., & Doeller, C. F. (2013). Mnemonic networks in the hippocampal formation: From spatial maps to temporal and conceptual codes. *Journal of Experimental Psychology: General*, *142*(4), 1231.

Miller, C. A., Campbell, S. L., & Sweatt, J. D. (2008). DNA methylation and histone acetylation work in concert to regulate memory formation and synaptic plasticity. *Neurobiology of learning and memory*, *89*(4), 599-603.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*,*298*(5594), 824-827.

Minatohara, K., Akiyoshi, M., & Okuno, H. (2016). Role of immediate-early genes in synaptic plasticity and neuronal ensembles underlying the memory trace. *Frontiers in molecular neuroscience*, *8*, 78.

Minichiello, L., Calella, A. M., Medina, D. L., Bonhoeffer, T., Klein, R., & Korte, M. (2002). Mechanism of TrkB-mediated hippocampal long-term potentiation. *Neuron*, *36*(1), 121-137.

Miranti, C. K., Ginty, D. D., Huang, G., Chatila, T., & Greenberg, M. E. (1995). Calcium activates serum response factor-dependent transcription by a Ras-and Elk-1-independent mechanism that involves a Ca2+/calmodulin-dependent kinase. *Molecular and cellular biology*, *15*(7), 3672-3684.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural brain research*, *6*(1), 57-77.

Miyashita, T., Kubik, S., Lewandowski, G., & Guzowski, J. F. (2008). Networks of neurons, networks of genes: an integrated view of memory consolidation.*Neurobiology of learning and memory*, *89*(3), 269-284.

Mizuno, K., & Giese, K. P. (2005). Hippocampus-dependent memory formation: do memory type-specific mechanisms exist?. *Journal of pharmacological sciences*, *98*(3), 191-197.

Mizuno, T., Kanazawa, I., & Sakurai, M. (2001). Differential induction of LTP and LTD is not determined solely by instantaneous calcium concentration: an essential involvement of a temporal factor. *European Journal of Neuroscience*, *14*(4), 701-708.

Mohorianu, I., Bretman, A., Smith, D. T., Fowler, E. K., Dalmay, T., & Chapman, T. (2017). Comparison of alternative approaches for analysing multi-level RNA-seq data. *PloS one*, *12*(8), e0182694.

Moita, M. A., Rosis, S., Zhou, Y., LeDoux, J. E., & Blair, H. T. (2003). Hippocampal place cells acquire location-specific responses to the conditioned stimulus during auditory fear conditioning. *Neuron*, *37*(3), 485-497.

Montero, V. M., & Jian, S. (1995). Induction of c-fos protein by patterned visual stimulation in central visual pathways of the rat. *Brain research*, *690*(2), 189-199.

Morgan, J. I., Cohen, D. R., Hempstead, J. L., & Curran, T. (1987). Mapping patterns of c-fos expression in the central nervous system after seizure. *Science*, *237*(4811), 192-197.

Morgulis A., Coulouris G., Raytselis Y., Madden T.L., Agarwala R., & Schäffer A.A. (2008) "Database indexing for production MegaBLAST searches." *Bioinformatics 15*:1757-1764.

Morishita, W., Connor, J. H., Xia, H., Quinlan, E. M., Shenolikar, S., & Malenka, R. C. (2001). Regulation of synaptic strength by protein phosphatase 1. *Neuron*, *32*(6), 1133-1148.

Morris, R. G., & Frey, U. (1997). Hippocampal synaptic plasticity: role in spatial learning or the automatic recording of attended experience?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*(1360), 1489-1503.

Morris, R. G. M., Hagan, J. J., & Rawlins, J. N. P. (1986). Allocentric spatial learning by hippocampectomised rats: a further test of the “spatial mapping” and “working memory” theories of hippocampal function. *The Quarterly journal of experimental psychology*, *38*(4), 365-395.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, *5*(7), 621.

Moser, M. B., & Moser, E. I. (1998). Functional differentiation in the hippocampus. *Hippocampus*, *8*(6), 608-619.

Moser, E., Moser, M. B., & Andersen, P. (1993). Spatial learning impairment parallels the magnitude of dorsal hippocampal lesions, but is hardly present following ventral lesions. *Journal of Neuroscience*, *13*(9), 3916-3925.

Mozzachiodi, R., & Byrne, J. H. (2010). More than synaptic plasticity: role of nonsynaptic plasticity in learning and memory. *Trends in neurosciences*, *33*(1), 17-26.

Mueller, O., Lightfoot, S., & Schroeder, A. (2016). RNA integrity number (RIN)–standardization of RNA quality control. *Agilent application note, publication*, 1-8.

Mulkey, R. M., Herron, C. E., & Malenka, R. C. (1993). An essential role for protein phosphatases in hippocampal long-term depression. *Science*, *261*(5124), 1051-1055.

Mullins, J. J., Peters, J., & Ganten, D. (1990). Fulminant hypertension in transgenic rats harbouring the mouse Ren-2 gene. *Nature*, *344*(6266), 541.

Mumby, D. G. (2001). Perspectives on object-recognition memory following hippocampal damage: lessons from studies in rats. *Behavioural brain research*, *127*(1-2), 159-181.

Mumby, D. G., Gaskin, S., Glenn, M. J., Schramek, T. E., & Lehmann, H. (2002). Hippocampal damage and exploratory preferences in rats: memory for objects, places, and contexts. *Learning & Memory*, *9*(2), 49-57.

Mumby, D. G., & Pinel, J. P. (1994). Rhinal cortex lesions and object recognition in rats. *Behavioral neuroscience*, *108*(1), 11.

Mundell, N. A., Beier, K. T., Pan, Y. A., Lapan, S. W., Göz Aytürk, D., Berezovskii, V. K., ... & Schier, A. F. (2015). Vesicular stomatitis virus enables gene transfer and transsynaptic tracing in a wide range of organisms. *Journal of Comparative Neurology*, *523*(11), 1639-1663.

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(6797), 722.

Nakamura, T., Nakamura, K., Lasser-Ross, N., Barbara, J. G., Sandler, V. M., & Ross, W. N. (2000). Inositol 1, 4, 5-trisphosphate (IP3)-mediated Ca2+ release evoked by metabotropic agonists and backpropagating action potentials in hippocampal CA1 pyramidal neurons. *Journal of Neuroscience*, *20*(22), 8365-8376.

Naldini, L., Blömer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F. H., ... & Trono, D. (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science*, *272*(5259), 263-267.

Nassi, J. J., Cepko, C. L., Born, R. T., & Beier, K. T. (2015). Neuroanatomy goes viral!. *Frontiers in neuroanatomy*, *9*, 80.

Nectow, A. R., Moya, M. V., Ekstrand, M. I., Mousa, A., McGuire, K. L., Sferrazza, C. E., ... & Friedman, J. M. (2017). Rapid molecular profiling of defined cell types using viral TRAP. *Cell reports*, *19*(3), 655-667.

Nedergaard, M., Ransom, B., & Goldman, S. A. (2003). New roles for astrocytes: redefining the functional architecture of the brain. *Trends in neurosciences*, *26*(10), 523-530.

Nevian, T., & Sakmann, B. (2006). Spine Ca2+ signaling in spike-timing-dependent plasticity. *Journal of Neuroscience*, *26*(43), 11001-11013.

Niehrs, C., & Pollet, N. (1999). Synexpression groups in eukaryotes. *Nature*, *402*(6761), 483.

Nikolaev, E., Werka, T., & Kaczmarek, L. (1992). C-fos protooncogene expression in rat brain after long-term training of two-way active avoidance reaction. *Behavioural brain research*, *48*(1), 91-94.

Nonaka, M., Kim, R., Sharry, S., Matsushima, A., Takemoto-Kimura, S., & Bito, H. (2014). Towards a better understanding of cognitive behaviors regulated by gene expression downstream of activity-dependent transcription factors. *Neurobiology of learning and memory*, *115*, 21-29.

Nygaard, V., Rødland, E. A., & Hovig, E. (2016). Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics*, *17*(1), 29-39.

Odajima, J., Wills, Z. P., Ndassa, Y. M., Terunuma, M., Kretschmannova, K., Deeb, T. Z., ... & Das, M. (2011). Cyclin E constrains Cdk5 activity to regulate synaptic plasticity and memory formation. *Developmental cell*, *21*(4), 655-668.

Ognjanovski, N., Schaeffer, S., Wu, J., Mofakham, S., Maruyama, D., Zochowski, M., & Aton, S. J. (2017). Parvalbumin-expressing interneurons coordinate hippocampal network dynamics required for memory consolidation. *Nature communications*, *8*, 15039.

Ohsawa, I., Nishimaki, K., Murakami, Y., Suzuki, Y., Ishikawa, M., & Ohta, S. (2008). Age-dependent neurodegeneration accompanying memory loss in transgenic mice defective in mitochondrial aldehyde dehydrogenase 2 activity. *Journal of Neuroscience*, *28*(24), 6239-6249.

Okuno, H. (2011). Regulation and function of immediate-early genes in the brain: beyond neuronal activity markers. *Neuroscience research*, *69*(3), 175-186.

Olarte-Sánchez, C. M., Kinnavane, L., Amin, E., & Aggleton, J. P. (2014). Contrasting networks for recognition memory and recency memory revealed by immediate-early gene imaging in the rat. *Behavioral neuroscience*, *128*(4), 504-522.

Oliet, S. H., Malenka, R. C., & Nicoll, R. A. (1997). Two distinct forms of long-term depression coexist in CA1 hippocampal pyramidal cells. *Neuron*, *18*(6), 969-982.

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental neurology*, *51*(1), 78-109.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain research*.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.

O'Neil, D., Glowatz, H., & Schlumpberger, M. (2013). Ribosomal RNA Depletion for Efficient Use of RNA-Seq Capacity. *Current protocols in molecular biology*, 4-19.

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in cognitive sciences*, *6*(12), 505-510.

Oshlack, A., & Wakefield, M. J. (2009). Transcript length bias in RNA-seq data confounds systems biology. *Biology direct*, *4*(1), 14.

Palmer, M. J., Isaac, J. T., & Collingridge, G. L. (2004). Multiple, developmentally regulated expression mechanisms of long-term potentiation at CA1 synapses. *Journal of Neuroscience*, *24*(21), 4903-4911.

Pandis, C., Sotiriou, E., Kouvaras, E., Asprodini, E., Papatheodoropoulos, C., & Angelatou, F. (2006). Differential expression of NMDA and AMPA receptor subunits in rat dorsal and ventral hippocampus. *Neuroscience*, *140*(1), 163-175.

Papa, M., Pellicano, M. P., Welzl, H., & Sadile, A. G. (1993). Distributed changes in c-Fos and c-Jun immunoreactivity in the rat brain associated with arousal and habituation to novelty. *Brain research bulletin*, *32*(5), 509-515.

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., & Hellmann, I. (2016). The impact of amplification on differential expression analyses by RNA-seq. *Scientific reports*, *6*, 25533.

Park, M., Penick, E. C., Edwards, J. G., Kauer, J. A., & Ehlers, M. D. (2004). Recycling endosomes supply AMPA receptors for LTP. *Science*, *305*(5692), 1972-1975.

Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, *321*(5894), 1322-1327.

Paxinos, G. & Watson, C. (1998). *The rat brain in stereotaxic co-ordinates (4th ed,).* New York: Academic Press.

Pedreira, M. E., Pérez-Cuesta, L. M., & Maldonado, H. (2004). Mismatch between what is expected and what actually occurs triggers memory reconsolidation or extinction. *Learning & Memory*, *11*(5), 579-585.

Pei, Q., Burnet, P. J., & Zetterström, T. S. (1998). Changes in mRNA abundance of microtubule-associated proteins in the rat brain following electroconvulsive shock. *Neuroreport*, *9*(3), 391-394.

Peixoto, L., Risso, D., Poplawski, S. G., Wimmer, M. E., Speed, T. P., Wood, M. A., & Abel, T. (2015b). How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets. *Nucleic acids research*, *43*(16), 7664-7674.

Peixoto, L. L., Wimmer, M. E., Poplawski, S. G., Tudor, J. C., Kenworthy, C. A., Liu, S., ... & Abel, T. (2015a). Memory acquisition and retrieval impact different epigenetic processes that regulate gene expression. *BMC genomics*, *16*(5), S5.

Pereira, S. R., Vasconcelos, V. M., & Antunes, A. (2011). The phosphoprotein phosphatase family of Ser/Thr phosphatases as principal targets of naturally occurring toxins. *Critical reviews in toxicology*, *41*(2), 83-110.

Perry, R. B. T., & Fainzilber, M. (2009, July). Nuclear transport factors in neuronal function. In *Seminars in cell & developmental biology* (Vol. 20, No. 5, pp. 600-606). Academic Press.

Pfarr, D. S., Rieser, L. A., Woychik, R. P., Rottman, F. M., Rosenberg, M., & Reff, M. E. (1986). Differential effects of polyadenylation regions on gene expression in mammalian cells. *DNA (Mary Ann Liebert, Inc.)*, *5*(2), 115-122.

Phillips, R. G., & LeDoux, J. E. (1994). Lesions of the dorsal hippocampal formation interfere with background but not foreground contextual fear conditioning. *Learning & Memory*, *1*(1), 34-44.

Pihlajamäki, M., Tanila, H., Könönen, M., Hänninen, T., Hämäläinen, A., Soininen, H., & Aronen, H. J. (2004). Visual presentation of novel objects and new spatial arrangements of objects differentially activates the medial temporal lobe subareas in humans. *European Journal of Neuroscience*, *19*(7), 1939-1949.

Pillay, S., Meyer, N. L., Puschnik, A. S., Davulcu, O., Diep, J., Ishikawa, Y. A., ... & Carette, J. E. (2016). An essential receptor for adeno-associated virus infection. *Nature*, *530*(7588), 108.

Pistoni, M., Verrecchia, A., Doni, M., Guccione, E., & Amati, B. (2010). Chromatin association and regulation of rDNA transcription by the Ras-family protein RasL11a. *The EMBO journal*, *29*(7), 1215-1224.

Pizarro, J. M., Haro, L. S., & Barea-Rodriguez, E. J. (2003). Learning associated increase in heat shock cognate 70 mRNA and protein expression. *Neurobiology of learning and memory*, *79*(2), 142-151.

Plath, N., Ohana, O., Dammermann, B., Errington, M. L., Schmitz, D., Gross, C., ... & Kobalz, U. (2006). Arc/Arg3. 1 is essential for the consolidation of synaptic plasticity and memories. *Neuron*, *52*(3), 437-444.

Poggio, G. F., Motter, B. C., Squatrito, S., & Trotter, Y. (1985). Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision research*, *25*(3), 397-406.

Poplawski, S. G., Schoch, H., Wimmer, M. E., Hawk, J. D., Walsh, J. L., Giese, K. P., & Abel, T. (2014). Object-location training elicits an overlapping but temporally distinct transcriptional profile from contextual fear conditioning. *Neurobiology of learning and memory*, *116*, 90-95.

Pozo, K., & Goda, Y. (2010). Unraveling mechanisms of homeostatic synaptic plasticity. *Neuron*, *66*(3), 337-351.

Puighermanal, E., Biever, A., Pascoli, V., Melser, S., Pratlong, M., Cutando, L., ... & Marsicano, G. (2017). Ribosomal protein S6 phosphorylation is involved in novelty-induced locomotion, synaptic plasticity and mRNA translation. *Frontiers in molecular neuroscience*, *10*.

Quirk, G. J., Muller, R. U., & Kubie, J. L. (1990). The firing of hippocampal place cells in the dark depends on the rat's recent experience. *Journal of Neuroscience*, *10*(6), 2008-2017.

Quirk, G. J., Muller, R. U., Kubie, J. L., & Ranck, J. B. (1992). The positional firing properties of medial entorhinal neurons: description and comparison with hippocampal place cells. *Journal of Neuroscience*, *12*(5), 1945-1963.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Racaniello, M., Cardinale, A., Mollinari, C., D'Antuono, M., De Chiara, G., Tancredi, V., & Merlo, D. (2010). Phosphorylation changes of CaMKii, ERK1/2, PKB/Akt kinases and CREB activation during early long-term potentiation at Schaffer collateral-CA1 mouse hippocampal synapses. *Neurochemical research*, *35*(2), 239-246.

Raj, A., & van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, *135*(2), 216-226.

Ramamoorthi, K., Fropf, R., Belfort, G. M., Fitzmaurice, H. L., McKinney, R. M., Neve, R. L., ... & Lin, Y. (2011). Npas4 regulates a transcriptional program in CA3 required for contextual memory formation. *Science*, *334*(6063), 1669-1675.

Ramírez-Amaya, V., Vazdarjanova, A., Mikhael, D., Rosi, S., Worley, P. F., & Barnes, C. A. (2005). Spatial exploration-induced Arc mRNA and protein expression: evidence for selective, network-specific reactivation. *Journal of Neuroscience*, *25*(7), 1761-1768.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., ... & Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome biology*, *14*(9), 3158.

Raymond, C. R. (2007). LTP forms 1, 2 and 3: different mechanisms for the 'long' in long-term potentiation. *Trends in neurosciences*, *30*(4), 167-175.

Raymond, C. R., & Redman, S. J. (2002). Different calcium sources are narrowly tuned to the induction of different forms of LTP. *Journal of Neurophysiology*, *88*(1), 249-255.

Raymond, C. R., Thompson, V. L., Tate, W. P., & Abraham, W. C. (2000). Metabotropic glutamate receptors trigger homosynaptic protein synthesis to prolong long-term potentiation. *Journal of Neuroscience*, *20*(3), 969-976.

Redondo, R. L., & Morris, R. G. (2011). Making memories last: the synaptic tagging and capture hypothesis. *Nature Reviews Neuroscience*, *12*(1), 17.

Redpath, N. T., Price, N. T., Severinov, K. V., & Proud, C. G. (1993). Regulation of elongation factor-2 by multisite phosphorylation. *European journal of biochemistry*, *213*(2), 689-699.

Reijmers, L. G., Perkins, B. L., Matsuo, N., & Mayford, M. (2007). Localization of a stable neural correlate of associative memory. *Science*, *317*(5842), 1230-1233.

Reimsnider, S., Manfredsson, F. P., Muzyczka, N., & Mandel, R. J. (2007). Time course of transgene expression after intrastriatal pseudotyped rAAV2/1, rAAV2/2, rAAV2/5, and rAAV2/8 transduction in the rat. *Molecular Therapy*, *15*(8), 1504-1511.

Reinhard, J. R., Kriz, A., Galic, M., Angliker, N., Rajalu, M., Vogt, K. E., & Ruegg, M. A. (2016). The calcium sensor Copine-6 regulates spine structural plasticity and learning and memory. *Nature communications*, *7*, 11613.

Renart, A., Song, P., & Wang, X. J. (2003). Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron*, *38*(3), 473-485.

Renelt, M., und Halbach, V. V. B., & und Halbach, O. V. B. (2014). Distribution of PCP4 protein in the forebrain of adult mice. *Acta histochemica*, *116*(6), 1056-1061.

Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychological review*, *74*(1), 71.

Rho, J. M., & Storey, T. W. (2001). Molecular ontogeny of major neurotransmitter receptor systems in the mammalian central nervous system: norepinephrine, dopamine, serotonin, acetylcholine, and glycine. *Journal of child neurology*, *16*(4), 271-280.

Rhodes, D. R., Kalyana-Sundaram, S., Tomlins, S. A., Mahavisno, V., Kasper, N., Varambally, R., ... & Chinnaiyan, A. M. (2007). Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia*, *9*(5), IN1-IN9.

Richter, J. D., & Klann, E. (2009). Making synaptic plasticity and memory last: mechanisms of translational regulation. *Genes & development*, *23*(1), 1-11.

Richter, J. D., & Sonenberg, N. (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature*, *433*(7025), 477.

Rickmann, M., & Wolff, J. R. (1995). S100 protein expression in subpopulations of neurons of rat brain. *Neuroscience*, *67*(4), 977-991.

Riedel, G., Micheau, J., Lam, A. G. M., Roloff, E. V. L., Martin, S. J., Bridge, H., ... & Morris, R. G. M. (1999). Reversible neural inactivation reveals hippocampal participation in several memory processes. *Nature neuroscience*, *2*(10), 898.

Risso, D., Schwartz, K., Sherlock, G., & Dudoit, S. (2011). GC-content normalization for RNA-Seq data. *BMC bioinformatics*, *12*(1), 480.

Risso, D., Ngai, J., Speed, T. P., & Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature biotechnology*, *32*(9), 896.

Rittié, L., & Perbal, B. (2008). Enzymes used in molecular biology: a useful guide. *Journal of cell communication and signaling*, *2*(1-2), 25-45.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139-140.

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, *11*(3), R25.

Robinson, M. D., & Smyth, G. K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, *23*(21), 2881-2887.

Robinson, M. D., & Smyth, G. K. (2008). Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*, *9*(2), 321-332.

Rochester, N., Holland, J., Haibt, L., & Duda, W. (1956). Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on information Theory*, *2*(3), 80-93.

Rogan, M. T., Leon, K. S., Perez, D. L., & Kandel, E. R. (2005). Distinct neural signatures for safety and danger in the amygdala and striatum of the mouse. *Neuron*, *46*(2), 309-320.

Rolls, E. T. (1996). A theory of hippocampal function in memory. *Hippocampus*, *6*(6), 601-620.

Rolls, E. T., Miyashita, Y., Cahusac, P. M., Kesner, R. P., Niki, H., Feigenbaum, J. D., & Bach, L. (1989). Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. *Journal of Neuroscience*, *9*(6), 1835-1845.

Romiguier, J., Ranwez, V., Douzery, E. J., & Galtier, N. (2010). Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. *Genome research*, *20*(8), 1001-1009.

Rosales, J., Han, B., & Lee, K. Y. (2003). Cdk7 functions as a cdk5 activating kinase in brain. *Cellular Physiology and Biochemistry*, *13*(5), 285-296.

Rosenberg, T., Gal-Ben-Ari, S., Dieterich, D. C., Kreutz, M. R., Ziv, N. E., Gundelfinger, E. D., & Rosenblum, K. (2014). The roles of protein expression in synaptic plasticity and memory consolidation. *Frontiers in molecular neuroscience*, *7*.

Rossato, J. I., Bevilaqua, L. R., Myskiw, J. C., Medina, J. H., Izquierdo, I., & Cammarota, M. (2007). On the role of hippocampal protein synthesis in the consolidation and reconsolidation of object recognition memory. *Learning & Memory*, *14*(1-2), 36-46.

Royo, N. C., Vandenberghe, L. H., Ma, J. Y., Hauspurg, A., Yu, L., Maronski, M., ... & Watson, D. J. (2008). Specific AAV serotypes stably transduce primary hippocampal and cortical cultures with high efficiency and low toxicity. *Brain research*, *1190*, 15-22.

Rudy, B., Fishell, G., Lee, S., & Hjerling-Leffler, J. (2011). Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Developmental neurobiology*, *71*(1), 45-61.

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current opinion in neurobiology*, *23*(2), 255-260.

Rumpel, S., LeDoux, J., Zador, A., & Malinow, R. (2005). Postsynaptic receptor trafficking underlying a form of associative learning. *Science*, *308*(5718), 83-88.

Saha, R. N., & Dudek, S. M. (2013). Splitting hares and tortoises: a classification of neuronal immediate early gene transcription based on poised RNA polymerase II. *Neuroscience*, *247*, 175-181.

Saha, R. N., Wissink, E. M., Bailey, E. R., Zhao, M., Fargo, D. C., Hwang, J. Y., ... & Dudek, S. M. (2011). Rapid activity-induced transcription of Arc and other IEGs relies on poised RNA polymerase II. *Nature neuroscience*, *14*(7), 848.

Salin, P. A., Malenka, R. C., & Nicoll, R. A. (1996). Cyclic AMP mediates a presynaptic form of LTP at cerebellar parallel fiber synapses. *Neuron*, *16*(4), 797-803.

Sanz, E., Yang, L., Su, T., Morris, D. R., McKnight, G. S., & Amieux, P. S. (2009). Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proceedings of the National Academy of Sciences*, *106*(33), 13939-13944.

Sato, E., Furuta, R. A., & Miyazawa, T. (2010). An endogenous murine leukemia viral genome contaminant in a commercial RT-PCR kit is amplified using standard primers for XMRV. *Retrovirology*, *7*(1), 110.

Sayols, S., Scherzinger, D., & Klein, H. (2016). dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC bioinformatics*, *17*(1), 428.

Scheyltjens, I., Laramée, M. E., Van den Haute, C., Gijsbers, R., Debyser, Z., Baekelandt, V., ... & Arckens, L. (2015). Evaluation of the expression pattern of rAAV2/1, 2/5, 2/7, 2/8, and 2/9 serotypes with different promoters in the mouse visual cortex. *Journal of Comparative Neurology*, *523*(14), 2019-2042.

Schratt, G. M., Tuebing, F., Nigh, E. A., Kane, C. G., Sabatini, M. E., Kiebler, M., & Greenberg, M. E. (2006). A brain-specific microRNA regulates dendritic spine development. *nature*, *439*(7074), 283.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., ... & Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC molecular biology*, *7*(1), 3.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., ... & Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, *473*(7347), 337.

Scott, H. L., Tamagnini, F., Narduzzo, K. E., Howarth, J. L., Lee, Y. B., Wong, L. F., ... & Uney, J. B. (2012). MicroRNA-132 regulates recognition memory and synaptic plasticity in the perirhinal cortex. *European Journal of Neuroscience*, *36*(7), 2941-2948.

Sefton AJ, Dreher B, Harvey AR. (2004). Visual system. In: G. Paxinos, editor. The rat nervous system. A handbook for neuroscientists. 3rd ed. Sydney: Academic Press; p. 1083-1165.

Serizawa, H., Tsuchihashi, Z., & Mizumoto, K. (1997). The RNA polymerase II preinitiation complex formed in the presence of ATP. *Nucleic acids research*, *25*(20), 4079-4084.

Seubert, P., Larson, J., Oliver, M., Jung, M. W., Baudry, M., & Lynch, G. (1988). Stimulation of NMDA receptors induces proteolysis of spectrin in hippocampus. *Brain research*, *460*(1), 189-194.

Shamir, M., Bar-On, Y., Phillips, R., & Milo, R. (2016). Snapshot: timescales in cell biology. *Cell*, *164*(6), 1302-1302.

Shanker, S., Paulson, A., Edenberg, H. J., Peak, A., Perera, A., Alekseyev, Y. O., ... & Grove, D. (2015). Evaluation of commercially available RNA amplification kits for RNA sequencing using very low input amounts of total RNA. *Journal of biomolecular techniques: JBT*, *26*(1), 4.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, *13*(11), 2498-2504.

Shatkin, A. J. (1985). mRNA cap binding proteins: essential factors for initiating translation. *Cell*, *40*(2), 223-224.

Sheng, M., & Greenberg, M. E. (1990). The regulation and function of c-fos and other immediate early genes in the nervous system. *Neuron*, *4*(4), 477-485.

Sheng, M., & Hoogenraad, C. C. (2007). The postsynaptic architecture of excitatory synapses: a more quantitative view. *Annu. Rev. Biochem.*, *76*, 823-847.

Sheng, M., Thompson, M. A., & Greenberg, M. E. (1991). CREB: a Ca (2+)-regulated transcription factor phosphorylated by calmodulin-dependent kinases. *Science*, *252*(5011), 1427-1430.

Shepherd, J. D., Rumbaugh, G., Wu, J., Chowdhury, S., Plath, N., Kuhl, D., ... & Worley, P. F. (2006). Arc/Arg3. 1 mediates homeostatic synaptic scaling of AMPA receptors. *Neuron*, *52*(3), 475-484.

Shepherd, A. J., Wilson, N. J., & Smith, K. T. (2003). Characterisation of endogenous retrovirus in rodent cell lines used for production of biologicals. *Biologicals*, *31*(4), 251-260.

Shevtsova, Z., Malik, J. M. I., Michel, U., Bähr, M., & Kügler, S. (2005). Promoters and serotypes: targeting of adeno-associated virus vectors for gene transfer in the rat central nervous system in vitro and in vivo. *Experimental physiology*, *90*(1), 53-59.

Shimizu, E., Tang, Y. P., Rampon, C., & Tsien, J. Z. (2000). NMDA receptor-dependent synaptic reinforcement as a crucial process for memory consolidation. *Science*, *290*(5494), 1170-1174.

Shimotohno, K., & Temin, H. M. (1981). Formation of infectious progeny virus after insertion of herpes simplex thymidine kinase gene into DNA of an avian retrovirus. *Cell*, *26*(1), 67-77.

Shires, K. L., & Aggleton, J. P. (2008). Mapping immediate-early gene activity in the rat after place learning in a water-maze: the importance of matched control conditions. *European Journal of Neuroscience*, *28*(5), 982-996.

Shuai, Y., Lu, B., Hu, Y., Wang, L., Sun, K., & Zhong, Y. (2010). Forgetting is regulated through Rac activity in Drosophila. *Cell*, *140*(4), 579-589.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., & Thompson, J. D. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, *7*(1), 539.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and social psychology review*, *4*(2), 108-131.

Smith, R. L., Traul, D. L., Schaack, J., Clayton, G. H., Staley, K. J., & Wilcox, C. L. (2000). Characterization of promoter function and cell-type-specific expression from viral vectors in the nervous system. *Journal of virology*,*74*(23), 11254-11261.

Smolen, P., Baxter, D. A., & Byrne, J. H. (2006). A model of the roles of essential kinases in the induction and expression of late long-term potentiation. *Biophysical journal*, *90*(8), 2760-2775.

Snyder, E. M., Philpot, B. D., Huber, K. M., Dong, X., Fallon, J. R., & Bear, M. F. (2001). Internalization of ionotropic glutamate receptors in response to mGluR activation. *Nature neuroscience*, *4*(11), 1079.

Sokolov, E.N., & Vinograda, O.S. (1975). *Neuronal Mechanisms of the Orienting Reflex.* Erlbaum, Mahwah, NJ.

Sokolova, O. O., Shtark, M. B., Lisachev, P. D., Pustylnyak, V. O., Pan, I. R., & Epstein, O. I. (2009). Expression of S100B and S100A6 genes during long-term posttetanic potentiation in the hippocampus. *Bulletin of experimental biology and medicine*, *148*(2), 227-229.

Soneson, C., & Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC bioinformatics*, *14*(1), 91.

Spies, D., & Ciaudo, C. (2015). Dynamics in transcriptomics: advancements in RNA-seq time course and downstream analysis. *Computational and structural biotechnology journal*, *13*, 469-477.

Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological review*, *99*(2), 195.

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current opinion in neurobiology*, *5*(2), 169-177.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380-1386.

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207-222.

Stark, C. E., & Okado, Y. (2003). Making memories without trying: medial temporal lobe activity associated with incidental memory formation during recognition. *Journal of Neuroscience*, *23*(17), 6748-6753.

Stellwagen, D., & Malenka, R. C. (2006). Synaptic scaling mediated by glial TNF-α. *Nature*, *440*(7087), 1054.

Steward, O., & Levy, W. B. (1982). Preferential localization of polyribosomes under the base of dendritic spines in granule cells of the dentate gyrus. *Journal of Neuroscience*, *2*(3), 284-291.

Steward, O., & Worley, P. F. (2001). Selective targeting of newly synthesized Arc mRNA to active synapses requires NMDA receptor activation. *Neuron*, *30*(1), 227-240.

Stranahan, A. M., Norman, E. D., Lee, K., Cutler, R. G., Telljohann, R. S., Egan, J. M., & Mattson, M. P. (2008). Diet-induced insulin resistance impairs hippocampal synaptic plasticity and cognition in middle-aged rats. *Hippocampus*, *18*(11), 1085-1088.

Su, J., Gorse, K., Ramirez, F., & Fox, M. A. (2010). Collagen XIX is expressed by interneurons and contributes to the formation of hippocampal synapses. *Journal of Comparative Neurology*, *518*(2), 229-253.

Suberbielle, E., Sanchez, P. E., Kravitz, A. V., Wang, X., Ho, K., Eilertson, K., ... & Mucke, L. (2013). Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid-β. *Nature neuroscience*, *16*(5), 613.

Sultan, F. A. (2013). Dissection of Different Areas from Mouse Hippocampus. *Bio-protocol* 3(21): e955.

Sun, X., & Lin, Y. (2016). Npas4: linking neuronal activity to memory. *Trends in neurosciences*, *39*(4), 264-275.

Sun, J., Nishiyama, T., Shimizu, K., & Kadota, K. (2013). TCC: an R package for comparing tag count data with robust normalization strategies. *BMC bioinformatics*, *14*(1), 219.

Sutton, M. A., & Carew, T. J. (2000). Parallel molecular pathways mediate expression of distinct forms of intermediate-term facilitation at tail sensory–motor synapses in Aplysia. *Neuron*, *26*(1), 219-231.

Sutton, M. A., & Schuman, E. M. (2006). Dendritic protein synthesis, synaptic plasticity, and memory. *Cell*, *127*(1), 49-58.

't Hoen, P. A. C., Friedländer, M. R., Almlöf, J., Sammeth, M., Pulyakhina, I., Anvar, S. Y., ... & van Ommen, G. J. B. (2013). Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nature biotechnology*, *31*(11), 1015.

Tavares, R. M., Mendelsohn, A., Grossman, Y., Williams, C. H., Shapiro, M., Trope, Y., & Schiller, D. (2015). A map for social navigation in the human brain. *Neuron*, *87*(1), 231-243.

Tagawa, Y., Kanold, P. O., Majdan, M., & Shatz, C. J. (2005). Multiple periods of functional ocular dominance plasticity in mouse visual cortex. *Nature neuroscience*, *8*(3), 380.

Takeda, K., Inoue, H., Tanizawa, Y., Matsuzaki, Y., Oba, J., Watanabe, Y., ... & Oka, Y. (2001). WFS1 (Wolfram syndrome 1) gene product: predominant subcellular localization to endoplasmic reticulum in cultured cells and neuronal expression in rat brain. *Human Molecular Genetics*, *10*(5), 477-484.

Takeichi, M., & Abe, K. (2005). Synaptic contact dynamics controlled by cadherin and catenins. *Trends in cell biology*, *15*(4), 216-221.

Tanaka, K. Z., Pevzner, A., Hamidi, A. B., Nakazawa, Y., Graham, J., & Wiltgen, B. J. (2014). Cortical representations are reinstated by the hippocampus during memory retrieval. *Neuron*, *84*(2), 347-354.

Tang, D., & Wang, J. H. (1996). Cyclin-dependent kinase 5 (Cdk5) and neuron-specific Cdk5 activators. In *Progress in cell cycle research* (pp. 205-216). Springer, Boston, MA.

Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., & Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. *Genome research*, *21*(12), 2213-2223.

Tashiro, A., & Yuste, R. (2004). Regulation of dendritic spine motility and stability by Rac1 and Rho kinase: evidence for two forms of spine motility. *Molecular and Cellular Neuroscience*, *26*(3), 429-440.

Taub, M. A., Bravo, H. C., & Irizarry, R. A. (2010). Overcoming bias and systematic errors in next generation sequencing data. *Genome medicine*, *2*(12), 87.

Taymans, J. M., Vandenberghe, L. H., Haute, C. V. D., Thiry, I., Deroose, C. M., Mortelmans, L., ... & Baekelandt, V. (2007). Comparative analysis of adeno-associated viral vector serotypes 1, 2, 5, 7, and 8 in mouse brain. *Human gene therapy*, *18*(3), 195-206.

Tebaldi, T., Re, A., Viero, G., Pegoretti, I., Passerini, A., Blanzieri, E., & Quattrone, A. (2012). Widespread uncoupling between transcriptome and translatome variations after a stimulus in mammalian cells. *BMC genomics*, *13*(1), 220.

Thiels, E., Norman, E. D., Barrionuevo, G., & Klann, E. (1998). Transient and persistent increases in protein phosphatase activity during long-term depression in the adult hippocampus in vivo. *Neuroscience*, *86*(4), 1023-1029.

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., ... & Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome research*, *13*(9), 2129-2141.

Thompson, K. R., Otis, K. O., Chen, D. Y., Zhao, Y., O'Dell, T. J., & Martin, K. C. (2004). Synapse to nucleus signaling during long-term synaptic plasticity: a role for the classical active nuclear import pathway. *Neuron*, *44*(6), 997-1009.

Thorsell, A., Michalkiewicz, M., Dumont, Y., Quirion, R., Caberlotto, L., Rimondini, R., ... & Heilig, M. (2000). Behavioral insensitivity to restraint stress, absent fear suppression of behavior and impaired spatial learning in transgenic rats with hippocampal neuropeptide Y overexpression. *Proceedings of the National Academy of Sciences*, *97*(23), 12852-12857.

Tian, X., Wang, G., Xu, Y., Wang, P., Chen, S., Yang, H., ... & Manyande, A. (2009). An improved tet-on system for gene expression in neurons delivered by a single lentiviral vector. *Human gene therapy*, *20*(2), 113-123.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411-423.

Tiedge, H., & Brosius, J. (1996). Translational machinery in dendrites of hippocampal neurons in culture. *The Journal of neuroscience*, *16*(22), 7171-7181.

Tischmeyer, W., & Grimm, R. (1999). Activation of immediate early genes and memory formation. *Cellular and Molecular Life Sciences CMLS*, *55*(4), 564-574.

Tiunova, A. A., Anokhin, K. V., & Rose, S. P. (1998). Two critical periods of protein and glycoprotein synthesis in memory consolidation for visual categorization learning in chicks. *Learning & memory*, *4*(5), 401-410.

Todorova, V., & Blokland, A. (2017). Mitochondria and synaptic plasticity in the mature and aging nervous system. *Current neuropharmacology*, *15*(1), 166-173.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55*(4), 189.

Tornow, S., & Mewes, H. W. (2003). Functional modules by relating protein interaction networks and gene expression. *Nucleic Acids Research*, *31*(21), 6283-6289.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols, 7(3), 562-578.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., ... & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*, *28*(5), 511.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*(3), 374-391.

Trinidad, J. C., Thalhammer, A., Burlingame, A. L., & Schoepfer, R. (2013). Activity-dependent protein dynamics define interconnected cores of co-regulated postsynaptic proteins. *Molecular & Cellular Proteomics*, *12*(1), 29-41.

Tsokas, P., & Blitzer, R. D. (2015). mTOR and the Regulation of Translational Capacity in Late Forms of Synaptic Plasticity. In *Synaptic Tagging and Capture* (pp. 99-132). Springer, New York, NY.

Tsokas, P., Grace, E. A., Chan, P., Ma, T., Sealfon, S. C., Iyengar, R., ... & Blitzer, R. D. (2005). Local protein synthesis mediates a rapid increase in dendritic elongation factor 1A after induction of late long-term potentiation. *Journal of Neuroscience*, *25*(24), 5833-5843.

Tuke, P. W., Tettmar, K. I., Tamuri, A., Stoye, J. P., & Tedder, R. S. (2011). PCR master mixes harbour murine DNA sequences. Caveat emptor!. *PloS one*, *6*(5), e19953.

Tullai, J. W., Schaffer, M. E., Mullenbrock, S., Sholder, G., Kasif, S., & Cooper, G. M. (2007). Immediate-early and delayed primary response genes are distinct in function and genomic architecture. *Journal of Biological Chemistry*.

Tuller, T., Veksler-Lublinsky, I., Gazit, N., Kupiec, M., Ruppin, E., & Ziv-Ukelson, M. (2011). Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome biology*, *12*(11), R110.

Turrigiano, G. (2007). Homeostatic signaling: the positive side of negative feedback. *Current opinion in neurobiology*, *17*(3), 318-324.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., & Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, *391*(6670), 892.

Tye, K. M., & Deisseroth, K. (2012). Optogenetic investigation of neural circuits underlying brain disease in animal models. *Nature Reviews Neuroscience*, *13*(4), 251.

Udagawa, T., Fujioka, Y., Tanaka, M., Honda, D., Yokoi, S., Riku, Y., ... & Katsuno, M. (2015). FUS regulates AMPA receptor function and FTLD/ALS-associated behaviour via GluA1 mRNA stabilization. *Nature communications*, *6*, 7098.

van den Brandt, J., Wang, D., Kwon, S. H., Heinkelein, M., & Reichardt, H. M. (2004). Lentivirally generated eGFP-transgenic rats allow efficient cell tracking in vivo. *Genesis*, *39*(2), 94-99.

Van der Perren, A., Toelen, J., Carlon, M., Van Den Haute, C., Coun, F., Heeman, B., ... & Baekelandt, V. (2011). Efficient and stable transduction of dopaminergic neurons in rat substantia nigra by rAAV 2/1, 2/2, 2/5, 2/6.2, 2/7, 2/8 and 2/9. *Gene therapy*, *18*(5), 517.

Van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in genetics*, *30*(9), 418-426.

van Hooijdonk, L. W., Ichwan, M., Dijkmans, T. F., Schouten, T. G., de Backer, M. W., Adan, R. A., ... & Fitzsimons, C. P. (2009). Lentivirus-mediated transgene delivery to the hippocampus reveals sub-field specific differences in expression. *BMC neuroscience*, *10*(1), 2.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, *277*(5324), 376-380.

Vavoulis, D. V., Francescatto, M., Heutink, P., & Gough, J. (2015). DGEclust: differential expression analysis of clustered count data. *Genome biology*, *16*(1), 39.

Vazdarjanova, A., & Guzowski, J. F. (2004). Differences in hippocampal neuronal population responses to modifications of an environmental context: evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles. *Journal of Neuroscience*, *24*(29), 6489-6496.

Vazdarjanova, A., Ramirez-Amaya, V., Insel, N., Plummer, T. K., Rosi, S., Chowdhury, S., ... & Barnes, C. A. (2006). Spatial exploration induces ARC, a plasticity-related immediate-early gene, only in calcium/calmodulin-dependent protein kinase II-positive principal excitatory and inhibitory neurons of the rat forebrain. *Journal of Comparative Neurology*, *498*(3), 317-329.

Vinogradova, O. S. (2001). Hippocampus as comparator: role of the two input and two output systems of the hippocampus in selection and registration of information. *Hippocampus*, *11*(5), 578-598.

Vo, N., Klein, M. E., Varlamova, O., Keller, D. M., Yamamoto, T., Goodman, R. H., & Impey, S. (2005). A cAMP-response element binding protein-induced microRNA regulates neuronal morphogenesis. *Proceedings of the National academy of Sciences of the United States of America*, *102*(45), 16426-16431.

Vogel, C., de Sousa Abreu, R., Ko, D., Le, S. Y., Shapiro, B. A., Burns, S. C., ... & Penalva, L. O. (2010). Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Molecular systems biology*, *6*(1), 400.

von Hertzen, L. S., & Giese, K. P. (2005). Memory reconsolidation engages only a subset of immediate-early genes induced during consolidation. *Journal of Neuroscience*, *25*(8), 1935-1942.

von Hungen, K., Mahler, H. R., & Moore, W. J. (1968). Turnover of protein and ribonucleic acid in synaptic subcellular fractions from rat brain. *Journal of Biological Chemistry*, *243*(7), 1415-1423.

von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., & Snel, B. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic acids research*, *31*(1), 258-261.

Wallace, D. J., Greenberg, D. S., Sawinski, J., Rulla, S., Notaro, G., & Kerr, J. N. (2013). Rats maintain an overhead binocular field at the expense of constant fusion. *Nature*, *498*(7452), 65.

Wan, H., Aggleton, J. P., & Brown, M. W. (1999). Different contributions of the hippocampus and perirhinal cortex to recognition memory. *The Journal of Neuroscience*, *19*(3), 1142-1148.

Wang, H., Ardiles, A. O., Yang, S., Tran, T., Posada-Duque, R., Valdivia, G., ... & Worley, P. (2016). Metabotropic glutamate receptors induce a form of LTP controlled by translation and arc signaling in the hippocampus. *Journal of Neuroscience*, *36*(5), 1723-1729.

Wang, D., Cui, Z., Zeng, Q., Kuang, H., Wang, L. P., Tsien, J. Z., & Cao, X. (2009). Genetic enhancement of memory and long-term potentiation but not CA1 long-term depression in NR2B transgenic rats. *PloS one*, *4*(10), e7486.

Wang, H. X., Gerkin, R. C., Nauen, D. W., & Bi, G. Q. (2005). Coactivation and timing-dependent integration of synaptic potentiation and depression. *Nature neuroscience*, *8*(2), 187.

Wang, S. H., & Morris, R. G. (2010). Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. *Annual review of psychology*, *61*, 49-79.

Wanisch, K., & Yáñez-Muñoz, R. J. (2009). Integration-deficient lentiviral vectors: a slow coming of age. *Molecular Therapy*, *17*(8), 1316-1332.

Warburton, E. C., & Brown, M. W. (2015). Neural circuitry for rat recognition memory. *Behavioural brain research*, *285*, 131-139.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., ... & Venables, B. (2016). gplots: various R programming tools for plotting data. R package version 3.0. 1. *The Comprehensive R Archive Network*.

Watson, D. J., Kobinger, G. P., Passini, M. A., Wilson, J. M., & Wolfe, J. H. (2002). Targeted transduction patterns in the mouse brain by lentivirus vectors pseudotyped with VSV, Ebola, Mokola, LCMV, or MuLV envelope proteins. *Molecular Therapy*, *5*(5), 528-537.

Watson, L. A., & Tsai, L. H. (2017). In the loop: how chromatin topology links genome structure to function in mechanisms underlying learning and memory. *Current opinion in neurobiology*, *43*, 48-55.

Waung, M. W., Pfeiffer, B. E., Nosyreva, E. D., Ronesi, J. A., & Huber, K. M. (2008). Rapid translation of Arc/Arg3. 1 selectively mediates mGluR-dependent LTD through persistent increases in AMPAR endocytosis rate. *Neuron*, *59*(1), 84-97.

Weeber, E. J., Levy, M., Sampson, M. J., Anflous, K., Armstrong, D. L., Brown, S. E., ... & Craigen, W. J. (2002). The role of mitochondrial porins and the permeability transition pore in learning and synaptic plasticity. *Journal of Biological Chemistry*, *277*(21), 18891-18897.

Wegner, A. M., Nebhan, C. A., Hu, L., Majumdar, D., Meier, K. M., Weaver, A. M., & Webb, D. J. (2008). N-wasp and the arp2/3 complex are critical regulators of actin in the development of dendritic spines and synapses. *Journal of Biological Chemistry*, *283*(23), 15912-15920.

Wei, Z., Wang, K., Qu, H. Q., Zhang, H., Bradfield, J., Kim, C., ... & Stanley, C. (2009). From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. *PLoS genetics*, *5*(10), e1000678.

Whitlock, J. R., Heynen, A. J., Shuler, M. G., & Bear, M. F. (2006). Learning induces long-term potentiation in the hippocampus. *science*, *313*(5790), 1093-1097.

Wiens, K. M., Lin, H., & Liao, D. (2005). Rac1 induces the clustering of AMPA receptors during spinogenesis. *Journal of Neuroscience*, *25*(46), 10627-10636.

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, *265*(5172), 676-679.

Winters, B. D., Forwood, S. E., Cowell, R. A., Saksida, L. M., & Bussey, T. J. (2004). Double dissociation between the effects of peri-postrhinal cortex and hippocampal lesions on tests of object recognition and spatial memory: heterogeneity of function within the temporal lobe. *Journal of Neuroscience*, *24*(26), 5901-5908.

Winters, B. D., & Reid, J. M. (2010). A distributed cortical representation underlies crossmodal object recognition in rats. *The Journal of Neuroscience*,*30*(18), 6253-6261.

Wittmann, M., Queisser, G., Eder, A., Wiegert, J. S., Bengtson, C. P., Hellwig, A., ... & Bading, H. (2009). Synaptic activity induces dramatic changes in the geometry of the cell nucleus: interplay between nuclear structure, histone H3 phosphorylation, and nuclear calcium signaling. *Journal of Neuroscience*, *29*(47), 14687-14700.

Wong, L. F., Azzouz, M., Walmsley, L. E., Askham, Z., Wilkes, F. J., Mitrophanous, K. A., ... & Mazarakis, N. D. (2004). Transduction patterns of pseudotyped lentiviral vectors in the nervous system. *Molecular Therapy*, *9*(1), 101-111.

Wong, L. F., Goodhead, L., Prat, C., Mitrophanous, K. A., Kingsman, S. M., & Mazarakis, N. D. (2006). Lentivirus-mediated gene transfer to the central nervous system: therapeutic and research applications. *Human gene therapy*, *17*(1), 1-9.

Wong, L. F., Ralph, G. S., Walmsley, L. E., Bienemann, A. S., Parham, S., Kingsman, S. M., ... & Mazarakis, N. D. (2005). Lentiviral-mediated delivery of Bcl-2 or GDNF protects against excitotoxicity in the rat hippocampus. *Molecular Therapy*, *11*(1), 89-95.

Worley, P. F., Bhat, R. V., Baraban, J. M., Erickson, C. A., McNaughton, B. L., & Barnes, C. A. (1993). Thresholds for synaptic activation of transcription factors in hippocampus: correlation with long-term enhancement. *Journal of Neuroscience*, *13*(11), 4776-4786.

Wu, H. H., Ivkovic, S., Murray, R. C., Jaramillo, S., Lyons, K. M., Johnson, J. E., & Calof, A. L. (2003). Autoregulation of neurogenesis by GDF11. *Neuron*, *37*(2), 197-207.

Wu, T. D., & Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, *26*(7), 873-881.

Xia, Z., & Storm, D. R. (2005). The role of calmodulin as a signal integrator for synaptic plasticity. *Nature Reviews Neuroscience*, *6*(4), 267.

Xiang, J. Z., & Brown, M. W. (1998). Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology*, *37*(4-5), 657-676.

Xiao, B., Tu, J. C., Petralia, R. S., Yuan, J. P., Doan, A., Breder, C. D., ... & Worley, P. F. (1998). Homer regulates the association of group 1 metabotropic glutamate receptors with multivalent complexes of homer-related, synaptic proteins. *Neuron*, *21*(4), 707-716.

Xu, R., Janson, C. G., Mastakov, M., Lawlor, P., Young, D., Mouravlev, A., ... & Leone, P. (2001). Quantitative comparison of expression with adeno-associated virus (AAV-2) brain-specific gene cassettes. *Gene therapy*, *8*(17), 1323.

Yeung, K. Y., Medvedovic, M., & Bumgarner, R. E. (2004). From co-expression to co-regulation: how many microarray experiments do we need?. *Genome biology*, *5*(7), R48.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1341.

Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., & Wang, S. (2010). GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*, *26*(7), 976-978.

Yudin, D., Hanz, S., Yoo, S., Iavnilovitch, E., Willis, D., Gradus, T., ... & Yoneda, Y. (2008). Localized regulation of axonal RanGTPase controls retrograde injury signaling in peripheral nerve. *Neuron*, *59*(2), 241-252.

Zagulska-Szymczak, S., Filipkowski, R. K., & Kaczmarek, L. (2001). Kainate-induced genes in the hippocampus: lessons from expression patterns. *Neurochemistry international*, *38*(6), 485-501.

Zangenehpour, S., & Chaudhuri, A. (2002). Differential induction and decay curves of c-fos and zif268 revealed through dual activity maps. *Molecular Brain Research*, *109*(1-2), 221-225.

Zenke, F., & Gerstner, W. (2017). Hebbian plasticity requires compensatory processes on multiple timescales. *Phil. Trans. R. Soc. B*, *372*(1715), 20160259.

Zenke, F., Hennequin, G., & Gerstner, W. (2013). Synaptic plasticity in neural networks needs homeostasis with a fast rate detector. *PLoS computational biology*, *9*(11), e1003330.

Zhang, Y. (2015). Network analysis reveals stage-specific changes in zebrafish embryo development using time course whole transcriptome profiling and prior biological knowledge. *BioData mining*, *8*(1), 26.

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*, *4*(1).

Zhang, X., Li, Q., Wang, L., Liu, Z. J., & Zhong, Y. (2016). Cdc42-dependent forgetting regulates repetition effect in prolonging memory retention. *Cell reports*, *16*(3), 817-825.

Zhang, B., Kirov, S., & Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic acids research*, *33*(suppl_2), W741-W748.

Zhang, X., Rosen, B. D., Tang, H., Krishnakumar, V., & Town, C. D. (2015). Polyribosomal RNA-Seq reveals the decreased complexity and diversity of the Arabidopsis translatome. *PloS one*, *10*(2), e0117699.

Zhang, S. J., Steijaert, M. N., Lau, D., Schütz, G., Delucinge-Vivier, C., Descombes, P., & Bading, H. (2007). Decoding NMDA receptor signaling: identification of genomic programs specifying neuronal survival and death. *Neuron*, *53*(4), 549-562.

Zhao, X., Lein, E. S., He, A., Smith, S. C., Aston, C., & Gage, F. H. (2001). Transcriptional profiling reveals strict boundaries between hippocampal subregions. *Journal of Comparative Neurology*, *441*(3), 187-196.

Zheng, H., Jia, H., Shankar, A., Heneine, W., & Switzer, W. M. (2011). Detection of murine leukemia virus or mouse DNA in commercial RT-PCR reagents and human DNAs. *PLoS One*, *6*(12), e29050.

Zhou, X., Lindsay, H., & Robinson, M. D. (2014). Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic acids research*, *42*(11), e91-e91.

Zhou, Z., Meng, Y., Asrar, S., Todorovski, Z., & Jia, Z. (2009). A critical role of Rho-kinase ROCK2 in the regulation of spine and synaptic function. *Neuropharmacology*, *56*(1), 81-89.

Zhu, X. O., Brown, M. W., McCabe, B. J., & Aggleton, J. P. (1995). Effects of the novelty or familiarity of visual stimuli on the expression of the immediate early gene c-fos in rat brain. *Neuroscience*, *69*(3), 821-829.

Zhu X. O., McCabe B. J., Aggleton J. P., & Brown M. W. (1996). Mapping visual recognition memory through expression of the immediate early gene c-fos. *NeuroReport* 7:1871–1875.

Ziff, E. B. (1997). Enlightening the postsynaptic density. *Neuron*, *19*(6), 1163-1174.

Ziv, Y., Burns, L. D., Cocker, E. D., Hamel, E. O., Ghosh, K. K., Kitch, L. J., ... & Schnitzer, M. J. (2013). Long-term dynamics of CA1 hippocampal place codes. *Nature neuroscience*, *16*(3), 264.

Zufferey, R., Donello, J. E., Trono, D., & Hope, T. J. (1999). Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *Journal of virology*, *73*(4), 2886-2892.

Zufferey, R., Dull, T., Mandel, R. J., Bukovsky, A., Quiroz, D., Naldini, L., & Trono, D. (1998). Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *Journal of virology*, *72*(12), 9873-9880.

Appendix 1: Solutions for TRAP

Dissection Buffer:

1X HBSS

2.5 mM HEPES-KOH, pH 7.4

35 mM Glucose

4 mM $NaHCO_3$

RNase-free water

100 µg/ml cycloheximide

(RNasin RNase inhibitor)

Homogenisation Buffer:

10 mM HEPES-KOH, pH 7.4

150 mM KCl

5 mM $MgCl_2$

RNase-free water

0.5 mM DTT

Protease Inhibitors (Complete EDTA-free)

RNasin RNase inhibitor (10x)

Superasin RNase inhibitor (10x)

100 µg/ml cycloheximide

0.15M KCl Wash Buffer:

10 mM HEPES-KOH, pH 7.4

150 mM KCl

5 mM $MgCl_2$

1% NP-40

RNase-free water

0.5 mM DTT

RNasin RNase inhibitor

100 µg/ml cycloheximide

0.35M KCl Wash Buffer:

10 mM HEPES-KOH, pH 7.4

150 mM KCl

5 mM $MgCl_2$

1% NP-40

RNase-free water

0.5 mM DTT

RNasin RNase inhibitor

100 µg/ml cycloheximide